www.kisa.or.kr

하반기

25년 사이버 위협 동향 및 26년 전망

과학기술정보통신부

KISA 한국인터넷진흥원

CONTENTS

하반기

25년 사이버 위협 동향 및 26년 전망

Trend / 사이버 위협 동향

Insights / 전문가 칼럼

하반기

25년 사이버 위협 동향 및 26년 전망

Part 1

Trend / 2025 하반기 사이버 위협 동향

Part. 1

01 침해사고 신고 현황

침해사고 신고 통계

과학기술정보통신부(한국인터넷진흥원)는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 '정보통신망법')」제48조의 3(침해사고 신고 등)에 따라 민간분야의 정보통신서비스 제공자로부터 침해사고 신고를 받고 있다. 연도별 침해사고 신고 통계를 살펴보면 2024년 1,887건에서 2025년 2,383건으로 전년 대비 26.3% 증가했다. 2023년부터 2025년까지 반기별 침해사고 신고 현황을 살펴보면 2023년 상반기 664건 하반기 613건, 2024년 상반기 899건 하반기 988건이며, 2025년 상반기 1,034건 하반기 1,349건의 침해사고 신고가 있었다. 2025년 상반기는 전년 동기대비 15% 증가한 반면, 하반기 침해사고 신고 건수는 1,349건으로 전년 동기대비 37% 증가했다.

표 1-1 | 유형별 침해사고 신고 현황

[단위 : 건수]

구분 \ 연도		2023년			2024년			2025년		
		上	下	종합	上	下	종합	上	下	종합
침해사고 신고	DDoS 공격	124(18.7%)	89(14.5%)	213(16.7%)	153(17.0%)	132(13.4%)	285(15.1%)	238(23.0%)	350(25.9%)	588(24.7%)
	악성코드	156(23.5%)	144(23.5%)	300(23.5%)	106(11.8%)	123(12.4%)	229(12.1%)	115(11.1%)	239(17.7%)	354(14.9%)
	(랜섬웨어)	(134, 0.2%)	(124, 20.2%)	(258, 20.2%)	(92, 10.2%)	(103, 10.4%)	(195, 10.3%)	(82, 7.9%)	(192, 14.2%)	(274, 11.5%)
	서버 해킹	320(48.2%)	263(42.9%)	583(45.6%)	504(56.1%)	553(56.0%)	1,057(56.0%)	531(51.4%)	522(38.7%)	1,053(44.2%)
	기타	64(9.6%)	117(19.1%)	181(14.1%)	136(15.1%)	180(18.2%)	316(16.7%)	150(14.5%)	238(17.6%)	388(16.3%)
합 계		664	613	1,277	899	988	1,887	1,034	1,349	2,383

과학기술정보통신부(한국인터넷진흥원)는 국내 민간분야 침해사고 신고 접수 시 DDoS 공격, 악성코드 감염, 서버 해킹 등으로 구분해 신고를 받고 있다.

2025년 유형별 침해사고 신고 통계를 살펴보면 디도스 공격(588건)이 전년(285건) 대비 약 2배로 급격히 증가했으며, 랜섬웨어 공격(274건)도 전년(195건) 대비 40.5%로 높게 증가했다. 전체 유형별 비중은 서버해킹이 44.2%로 가장 높았으며, 그 다음으로 DDoS 공격이 24.7%, 악성코드 감염이 14.9%로 나타났다.

2023년부터 2025년까지 반기별 침해사고 신고 현황을 살펴보면 웹셸 삽입, 서버 데이터 변조, 피싱, 해킹 경유지 악용 등 서버해킹이 2023년 상반기 320건, 하반기 263건, 2024년 상반기 504건, 하반기 553건, 2025년 상반기 531건, 하반기 522건으로 가장 많은 신고를 받은 것으로 나타났다. 그 다음으로는 DDoS 공격 신고가 2023년 상반기 124건, 하반기 89건, 2024년 상반기 153건, 하반기 132건, 2025년 상반기 238건, 하반기 350건으로 많았으며, 다음으로 악성코드 감염 신고가 2023년 상반기 156건, 하반기 144건, 2024년 상반기 106건,하반기 123건, 2025년 상반기 115건, 하반기 239건이었다. 정보 유출, 스팸 문자 발송 등 기타 신고 건은 2023년 상반기 64건, 하반기 117건, 2024년 상반기 136건, 하반기 180건, 2025년 상반기 150건, 하반기 238건으로 나타났다.

2025년 침해사고 신고 유형 중 악성코드 감염 통계를 살펴보면 악성코드 감염 비중 중 77% 이상을 랜섬웨어 신고가 차지하고 있었으며, 랜섬웨어로 인한 피해기업의 신고 건수는 2023년 상반기 134건, 하반기 124건, 2024년 상반기 92건, 하반기 103건, 2025년 상반기 82건, 하반기 192건인 것으로 나타났다.

규모별 랜섬웨어 침해사고 현황을 살펴보면 중견기업은 전년 대비 50% 증가한 51건, 중소기업은 전년 대비 29.3% 증가한 194건으로 나타났다.

랜섬웨어 침해사고와 관련해서 신고 기관(업)의 백업 여부 현황을 살펴보면 전체 백업률은 2023년 상반기 47%, 하반기 70.2%, 2024년 상반기 69.6%, 하반기 75.7% 2025년 상반기 76.8%, 하반기 78.6%로 백업 비중이 증가하는 추세를 보였다. 이 중 2023년 상반기 42.9%, 하반기 35.6%, 2024년 상반기 40.6%, 하반기 43.6%가 2025년 상반기 44.4%/ 하반기 23.2%가 백업까지 감염된 것으로 파악됐다.

전체 침해사고 중 랜섬웨어 감염 비중은 10~11% 수준이나, 국내 기업들의 피해 최소화 및 신속한 복구를 지원하기 위해서 과기정통부(KISA)는 보호나라(www.boho.or.kr) → 알림마당 → 보안공지에 랜섬웨어 대응을 위한 백업데이터 8대 수칙을 마련하여 공지했다.

2026년 사이버 위협 전망

1

AI활용 위협의 일상화,
AI 서비스에 대한 위협 증가

2

기술지원종료,
쉐도우 시스템에 대한 위협 증가

3

취약한 클라우드 보안
환경의 공격 증가

4

유출된 개인정보를 악용한
2차 사이버 위협 증가

2025년 사이버 위협 분석

1

국민 생활과 밀접한
다수 침해사고 사례

2

오픈소스 및 IoT 생태계를
악용한 공급망 사고사례

3

다수 고객망 운영기업 대상,
랜섬웨어 사고사례

AhnLab Genians IGLOO NSHC S2W PLAINBIT
SK 쉴더스 CISCO TALOS Google Microsoft TREND MICRO zscaler

2025년 사이버 위협 분석과 2026년 전망

2025년은 통신, 유통, 금융 등 국민생활 밀접분야에서 전례 없이 잦은 침해사고와 함께 개인정보 유출사고가 발생했다. 이로인해 많은 국민들이 개인정보 유출에 따른 2차 피해가 발생할 수 있다는 걱정 등 불안감을 겪었으며, 직접적인 금전 피해를 입기도 했다. 이와 같이 올해 국내에서 발생한 침해사고는 국민 일상 및 국가 경제에도 큰 영향을 미쳤다고 볼 수 있다. 국민의 일상 생활까지 침해사고에 영향을 받고 있는 상황에서 올 한 해 주요 사이버 침해사고 건들을 되돌아보며 사이버 보안 태세를 정비하고자 한다.

이를 위해 과학기술정보통신부(이하 '과기정통부')와 한국인터넷진흥원(이하 KISA)은 국내·외 사이버 위협 인텔리전스 네트워크*와 함께 올 한 해 발생했던 사이버 침해사고를 3가지 주제(국민생활, 공급망 보안, 랜섬웨어)를 중심으로 분류하여 분석했다. 또한, AI 확산, 사이버 보안 관련 기술 및 정책 변화 등을 고려해 '26년에 예상되는 사이버 위협 4가지 주제(AI, 자산관리, 클라우드, 개인침해)를 선정하였고, 이를 구체적으로 분석해 국내 정부·공공 기관 및 기업 등이 사이버 위협 대응에 참고할 수 있도록 제시했다.

* 과기정통부(KISA)와 국내·외 보안업체가 운영하는 사이버 위협 정보 공유 협력 네트워크로 (국내) 안랩, 지니언스, 이글루코퍼레이션, NSHC, S2W, SK쉴더스, 플레인비트, (해외) Cisco Talos, Google, Microsoft, Trend Micro, Zscaler 12개 기업이 참여 중

본 보고서를 통해 갈수록 고도화 지능화 되고 있는 사이버 위협에 체계적으로 대응할 수 있도록 내부 보안관리체계를 재정비하여 보안사각 지대를 최소화하고 선제적으로 대응할 수 있는 기반을 마련하는 계기가 될 수 있기를 기대한다.

Part. 1

02 2025년 사이버 위협 분석

1 일상을 위협하는 생활 밀접 인프라에 대한 침해사고

- 생활 밀접 분야에서 연속적인 침해사고 및 개인정보 유출 발생
- 침해사고 발생을 축소 · 은폐하려는 것으로 의심되는 사례 등 국민불안 심화

2025년은 2,383건으로 역대 가장 많은 사이버 침해사고('25.11월 기준 2,167건, '24년 1,887건)가 발생했다. 침해사고에 따른 개인정보 유출 규모도 역대급이라고 할 수 있으며, 침해사고 발생 분야도 통신, 유통, 금융 등 국민 생활 밀접 분야에서 발생하면서 많은 국민을 불안하게 했다.

주요 침해사고 사례

- 'SKT 해킹' 3년 전부터 시작...유심 정보 2,700만건 털렸다(4월)
- '랜섬웨어 해킹' 예스24, 두 달 만에 또 먹통(6월, 8월)
- SK텔레콤 이어 KT · 롯데카드까지...'해킹 포비아' 대한민국(8월)

SKT 침해사고는 외부 인터넷 연결 접점이 있는 시스템에 대한 해킹을 통해 내부 서버에 침투 후 장기간에 걸친 내부 시스템 탐색과 거점 확보를 거쳐 SKT 전체 고객의 유심정보를 탈취해가는 과정에서 탐지되어 그 전모가 드러났다. 이 사고를 통해 SKT 약 2천6백만 고객의 유심정보가 유출된 것으로 확인되면서 불안한 가입자들이 유심교체를 위해 대리점에 긴 줄을 서는 안타까운 상황이 벌어졌고, 많은 SKT 고객이 다른 통신사로 이동하는 사태도 발생했다.

국내 대표 온라인 서점 예스24는 최초에는 시스템 장애로 인해 서비스 마비가 발생한 것으로 발생했으나, 하루 뒤가 돼서야 랜섬웨어로 인한 장애로 발표하면서 침해사고 사실을 축소하려고 했던 것은 아닌가 하는 의심을 사게 됐다. 또한, 예스24의 발표와는 달리 KISA의 기술지원에 동의하지 않은 사실이 드러나면서 사회적 비판의 대상이 되기도 했다. 서비스 복구 이후 2달 만에 또다시 랜섬웨어로 인한 서비스 마비가 발생했다. 최초 사고 이후 수립한 복구체계를 통해 당일 내 서비스를 복구했으나, 연이은 사이버 침해사고로 언론을 통해 '해킹 맛집'이라는 오명을 얻게 됐다.

8월에는 美 사이버 보안 매거진 '프랙'을 통해 北 해킹조직(Kimsuky)가 KT와 LGU+를 해킹했다는 정보가 공개됐다. 미상의 화이트해커가 Kimsuky로 보이는 해킹 조직의 서버에 침투하여 그들이 해킹을 통해 보유 중인 KT의 인증서, LGU+ 서버 정보 및 직원 정보 등을 탈취했다는 내용이다. 이와 관련 과기정통부 민관합동조사단은 양 통신사에 대해서 면밀하게 조사하고 있으며, 향후 그 결과를 투명하게 공개할 예정이다.

수도권 일부 지역에서 KT 이동통신망 이용 고객의 휴대전화에서 이용한 적 없는 소액결제 사고가 발생, 과기정통부에서 민관합동조사단을 구성해 조사한 결과 불법 초소형 기지국을 이용한 무단 소액결제가 발생한 것을 확인했다. 민간합동조사단은 철저한 조사를 통해 지난 11월 6일 중간조사 결과를 발표했다. 또한 조사 과정에서 KT가 작년 BPFDoor 악성코드에 감염된 것을 당국에 신고 없이 자체 조치하고, 은폐한 사실도 확인되면서 또 다시 KT는 사회적 지탄을 받았다.

생활 인프라에 대한 사이버 공격은 우리나라뿐만 아니라 전세계적으로 발생하고 있다. 유럽 전역의 주요 공항들의 항공 탑승 시스템 해킹, 영국 대형 소매업체들에 대한 연쇄 해킹, 명품 브랜드 그룹 해킹 등이 발생했으며 미국에서도 지난해 9개의 통신사*에서 국가 배후 해킹조직으로 추정되는 해킹 조직에 의한 침해사고가 발생했다.

* AT&T, 버라이즌, T-Mobile, 루멘 테크놀로지스, 카터 커뮤니케이션즈, 컨솔리데이티드 커뮤니케이션즈, 윈드스트림 커뮤니케이션즈, 비아셋, 익명 통신사 등

2 오픈소스 및 저가형 IoT 생태계를 악용한 공급망 공격

- **오픈소스 저장소를 통해 악성코드를 유포하려는 시도 발생**
- **백도어 탑재, 보안 설정 미흡 등 보안에 취약한 형태로 유통되는 저가형 단말기 악용**

오픈소스 생태계를 통한 AI 공급망 공격도 본격화되고 있다. 美 보안기업 Sonatype에 따르면 2025년 2분기 신규 악성 패키지가 전년동기 대비 188% 급증한 16,279개가 등록되었으며, Hugging Face, GitHub, NPM 등 개발자들이 신뢰하는 오픈소스 플랫폼이 주요 공격 경로로 악용되고 있다. 특히 AI 모델 저장소와 전통적 소프트웨어 패키지 관리 시스템이 결합하면서 공격표면이 크게 확대되고 있다.

주요 사고 사례 및 동향

- FBI, BADBOX 2.0 악성코드로 100만 대 이상 감염 경고(6월)
- npm, PyPI 공급망 악성코드 공격 피해 심각...오픈소스 생태계 위협 급증(9월)
- 북한 해커, npm 악성 패키지 197개 추가..개발자 타깃 대규모 악성코드 공격 확인(12월)

국제적으로 오프소스 기반의 소프트웨어(SW) 공급망을 노린 사이버 공격이 전례 없이 급증하여 SW 개발 과정에서 오픈소스를 활용하는 전 세계 SW 개발자들이 심각한 위험에 노출되었다. 공격자가 GitHub 계정을 탈취하여 코드 저장소에 악성 코드가 삽입된 패키지를 추가하고, 이를 배포하도록 하여 수천 개의 보안 토큰을 유출하는 'Ghost Action'이 9월 발견됐다. 특히 최근에는 북한 해킹 조직이 NPM(Node Package Manager) 오픈소스 패키지 저장소에 무려 197개의 악성 패키지를 추가하며 악성코드를 유포하려고 시도한 공격이 발견되기도 했다.('25.12월)

AI 기반 코드 생성 또한 공급망 보안 취약점을 증폭시키고 있다. GitLab의 조사에 따르면 SW 개발과정에서 개발자의 AI 도구를 활용한 코드 생성이 크게 증가했으나, 보안 검증 프로세스는 이를 따라가지 못하고 있다. GitLab이 약 3천2백여명의 실무자를 대상으로 조사한 결과, 73%가 AI 생성 코드에서 보안취약점 등의 문제를 경험했으나 97%가 SW개발에 AI를 사용 중이거나 사용할 계획으로 나타나면서 보안부서가 느끼는 위기감이 개발부서의 생산성 향상에 가려져 있는 것으로 나타났다. 이러한 속도 중심의 SW 개발 문화는 SW 구성 요소 간 의존성 검증, 코드 리뷰, 보안 테스트를 생략하게 만들어 공격자가 의도한 공급망 공격의 성공 확률을 높이고 있다.

저가형 IoT 기기의 확산과 제조단계의 보안 미비로 인해 출하 전부터 오염된 장비가 대규모로 시장에 유입되는 사례도 발생했다. 2025년 상반기에 드러난 'BadBox 2.0' 사례는 중국계 제조사가 생산한 안드로이드 기반 셋톱박스·스마트TV·태블릿 등에 백도어가 선탑재된 채 유통된 사건으로, 출하 시점부터 이미 감염된 디바이스가 전 세계적으로 수십만 대 이상 유통됐다. 이들 기기는 사용자 동의 없이 광고사기, 프록시 중계, 계정탈취 등의 불법 행위에 활용됐다.

SW 공급망 보안을 위한 효과적인 수단으로 SBOM(Software Bill of Materials) 활용이 강조되고 있으며, 이를 통해 오픈소스에 내재되어 유통될 수 있는 보안취약점에 대한 추적관리와 정기적 재분석을 통해 SW 공급망에 대한 위험관리가 가능해질 것으로 보고 있다. 그리고 IoT 장비에 대해서 기본 비밀번호 변경 및 인증 강화, 인증된 펌웨어 사용 및 최신 업데이트 적용, 화이트리스 기반 통신 및 불필요 포트·서비스 비활성화 등 장비에 대한 지속적인 관리가 필요하다.

과기정통부는 지난 12월7일 개인정보위, 방미통위, 경찰청과 합동으로 「IP카메라 보안강화 방안('24년 11월)」을 발표했다. 정부는 이를 통해 지속되는 IP카메라 해킹과 영상 유출로 인한 피해 최소화와 국민 사생활 보호를 위해 해킹에 취약한 IP카메라 보안 조치, IP카메라 해킹 피해자 보호, 생활 밀접 시설의 IP카메라 보안인증 의무화 등을 추진할 계획이다.

3 랜섬웨어 공격 대상 확대와 기업-고객 연계 공격 강화

- 과거 민감분야(교육, 의료 등) 회피 경향과 달리 최근에는 분야 구분 없이 공격
- B2C 기업이나 단일 기업 침해를 통한 연쇄 패턴 공격 증가

사이버 공격의 주류가 된 랜섬웨어 그룹의 공격 대상이 연구·제조·에너지 분야를 넘어 교육·의료 등 공격 대상 분야를 계속 확대 해가고 있다. 그 수법 또한 AI 기술을 이용한 자동화나 연계형 공격으로 고도화 되어가고 있다.

주요 사고 사례 및 동향

- 美 신장치료 전문 기업 다비타, Interlock 랜섬웨어에 의한 침해 발생(4월)
- 美 병원 네트워크가 랜섬웨어 공격으로 전산 시스템 대거 중단(5월~6월)
- Qilin 랜섬웨어 그룹에 의한 다수의 국내 자산운용사 침해사고 발생(8월)
- 혈액공급기관 상대 초유의 랜섬웨어 공격.."정보를 쥐고 있으니 협상하자"(11월)

2025년 들어 과거 공격 대상에서 교육, 의료 분야를 제외하던 그룹들이 방향을 바꾸어, 교육기관과 헬스케어 기업까지 공격 범위를 넓혔다. 대표적인 피해 사례로는 DaVita, Texas Digestive Specialists, Naper Grove Vision Care, Clinical Diagnostics, Shamir Medical Center 등이 있으며, 헬스케어 분야에서만 120건 이상의 피해가 보고됐다. 한편, 미국에서는 의료기관을 겨냥한 공격이 증가하자 FBI, CISA, MS-ISAC 등 주요 보안 기관은 Medusa 및 Interlock 랜섬웨어 그룹 관련 공동 주의보를 발표하면서, 관련 기관에 강화된 보안 대응을 촉구했다.

헬스케어 이외에도 B2C 기업을 노린 공격이 다수 확인됐다. 공학 및 과학용 소프트웨어 MATLAB을 개발한 MathWorks는 랜섬웨어 공격으로 주요 플랫폼과 서비스가 일시 중단되어 엔지니어, 교수, 학생 등 사용자층이 직접적인 피해를 입었다. 국내에서도 도서 및 e-book, 공연 티켓 등을 제공하는 B2C 기업이 공격받아 서비스가 마비되는 사고가 발생했다. 이렇듯 B2C 기업이 공격받는 경우 기업 내부의 업무 마비뿐 아니라 서비스 이용 고객에게 직접적인 피해를 초래한다는 점에서 다른 기업들의 피해와 차별화된다. 또한 피해 기업은 대규모 고객 불만과 보상 요구로 인한 2차 피해와 평판 리스크까지 감수해야 하는 상황이 발생하고, 이로 인해 랜섬웨어 협상 과정에서 공격자가 우위를 점하기 쉬운 구도가 형성된다.

이러한 협상 구도가 형성되는 가운데, Qilin 랜섬웨어 그룹은 이러한 흐름을 적극적으로 활용했다. 2025년 2분기부터 활발히 활동한 이들은 국내 금융 IT 컨설팅 및 데이터 운영 기업을 공격해 대규모 데이터 유출을 일으켰다. 특히 피해 기업의 서버에 저장된 다수의 자산운용사 정보를 분석해, 이를 개별 피해 사례로 분리한 뒤 각각 협상을 시도하는 차별화된 공격 방식을 전개했다.

랜섬웨어 공격에 대비하기 위해 각 기업은 공격표면 관리(ASM, Attack Surface Management), 내부 시스템 보안 점검, 백업 등을 강화해야 한다. 최근에는 데이터 암호화와 동시에 탈취한 정보를 공개하겠다고 협박하는 이중 갈취(Double Extortion) 수법이 확산하면서, 백업의 중요성이 예전보다 낮게 인식되는 경향도 나타나고 있다. 하지만 실제로는 피해 발생 시 신속한 서비스 복구와 업무 지속을 위해 정기적인 백업과 복구 체계 구축이 여전히 필수적이다. 특히 공격자는 내부 시스템을 장악한 뒤 백업 서버나 복구용 스토리지까지 암호화하는 경우가 많기 때문에, 네트워크와 완전히 분리된 콜드 백업(Cold Backup) 환경을 구축해 안전한 복구 지점을 확보해야 하며, 주기적인 훈련을 통해 백업의 실효성을 강화해야 한다.

KISA는 랜섬웨어 대응 역량이 부족한 지역·중소·영세기업을 대상으로 무상 보안취약점 점검과 서버 보안점검(내서버돌보미)을 지원 중이며, 한국정보보호산업협회(KISIA)도 랜섬웨어 대응 보안솔루션 패키지 지원사업을 진행하고 있어, 랜섬웨어 예방 대책이 필요한 기업은 언제든지 도움을 받을 수 있다. 또한 랜섬웨어에 대비할 수 있는 백업 요령도 보호나라(www.boho.or.kr) → 보안공지를 통해서 안내한 바 있다.

Part. 1

03 2026년 사이버 위협 전망

1 AI 기반 사이버 위협 및 AI 서비스에 대한 사이버 공격 증가

- AI 기반으로 기존 방어체계를 우회하려는 사이버 공격 시도 증가
- AI 서비스 모델 자체를 공격 대상으로 삼는 기법도 본격화
- 공격 전 과정을 통합한 AI 기반 공격 체계를 통해 사이버 공격의 자동화 · 최적화

사이버 공격자들의 AI 활용이 본격화하면서 2026년에는 AI를 활용한 사이버 공격이 더욱 정교하고 다양화될 것으로 전망된다. 2026년에는 딥페이크 음성·영상 기반 피싱이 실시간 음성 통화 및 화상회의에까지 적용되어, 신뢰 기반 커뮤니케이션을 위협할 가능성이 높다. 또한 Gartner는 오는 2027년까지 AI 기술을 활용한 공격 도구가 사용자 계정 탈취에 소요되는 시간을 절반으로 단축할 것으로 예측했다.

주요 사고 사례 및 동향

- MS, 외부 데이터에 공격 지시를 심어 AI가 정보유출하게 하는 공격자 전략 발표(7월)
- 세계 최초 LLM 기반 악성코드 등장..인공지능 기술의 무기화 본격화(7월)
- 결국 등장한 AI 기반 랜섬웨어...보안업계 우려 현실화(8월)

AI가 악성 스크립트를 실시간으로 생성하고 실행하는 공격 방식이 개념 증명 수준을 넘어 실제 공격에 활용될 것으로 보인다. PromptLock 랜섬웨어, LAMEHUG 악성코드 등의 실제 사례를 통해 공격자가 피해자의 환경에 맞춰 즉석에서 악성코드를 생성하여 기존 탐지 시스템을 우회하는 형태로 진화했음을 확인할 수 있다. 이러한 변화는 기존 보안 체계의 한계를 드러내며, AI 악용 방지 기술, 모델 무결성 검증, 실시간 행위 기반 탐지 등 새로운 AI 위협 대응 전략의 필요성을 제기했다.

AI 서비스 모델 자체를 공격 대상으로 삼는 공격 기법도 본격화될 것으로 보인다. 공격자는 챗봇, 자동 분석 시스템, 보안 AI 등에 악의적인 내용을 주입하거나 학습 데이터를 조작함으로써 의도된 오작동이나 정보 노출을 유도할 수 있다. 이는 AI 기반 보안 솔루션의 신뢰성과 안정성을 근본적으로 흔드는 위협이 될 수 있다. 특히 기업이나 기관에서 자체 구축한 AI에 침투할 경우 기밀 정보 유출도 가능해 새로운 정보 유출 경로가 될 수 있다.

마지막으로 공격 전 과정을 통합한 AI 기반 공격 체계가 등장할 가능성이 높다. AI 기반 사이버 공격이 정찰, 침투, 취약점 탐지, 악성코드 개발, 침해 후 활동 등 모든 공격 단계에서 AI를 활용하여 더욱 빠르고 정교하며 광범위하게 진화하여 AI가 대상 선정, 취약점 탐색, 협상까지 스스로 수행하며 인간의 개입을 최소화하여 공격을 최적화·자동화할 것으로 예상된다.

AI를 악용한 사이버 공격은 기존 시그니처 기반 탐지 방식으로는 대응이 어려운 만큼, 비정상적인 코드 생성 행위를 식별할 수 있는 새로운 탐지 체계가 요구된다. 또한 딥페이크 음성 및 영상 기반 피싱에 대응하기 위해 실시간 인증 기술과 AI 기반 위조 탐지 알고리즘의 고도화가 필요하다.

생성형 AI의 악용 확산에 대응하기 위해, 기업과 기관은 AI 보안을 설계 단계(Security by Design)에서부터 내재화해야 한다. 학습 데이터의 출처·무결성을 검증하고, 프롬프트 인젝션·데이터 중독 등 내부 위협을 탐지할 수 있는 필터링 체계를 구축해야 한다. LLM 보안 게이트웨이(프롬프트 필터링·권한 최소화·로깅)를 도입해 입력·응답 과정을 상시 모니터링하고, 모델 접근 권한과 로그를 중앙에서 관리함으로써 의도치 않은 정보 노출을 방지해야 한다.

AI 공급망 전체에 대한 무결성 검증과 정기적인 레드팀 테스트를 통해, 악성 모델 주입이나 데이터 오염을 조기에 탐지할 수 있는 환경을 마련해야 한다. 마지막으로, AI 방어 자동화와 인간 분석가의 결합(Human-in-the-Loop Defense)이 필수적이다. AI 탐지 결과를 전문가가 검증·보완하는 상호보완형 구조를 통해 "AI로부터의 위협을 AI와 사람이 함께 방어하는 체계"를 구축해야 한다.

2 EOS와 방치된 미사용 시스템이 해킹 통로로 악용

- **EOS*와 패치 관리가 미흡한 시스템을 악용한 침해사고 반복**
  * EOS(End of Support): SW 개발사나 HW 제조사가 더 이상 기술지원을 하지 않는 제품
- **IT자산 파악, 레거시 시스템을 네트워크 격리, IT인프라 전반 패치 관리 강화 등 필요**

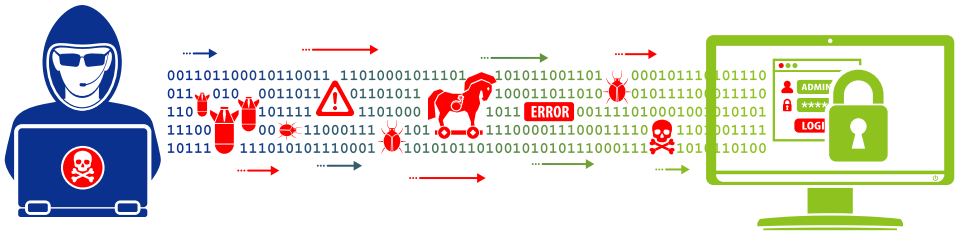
'관리의 빈틈'을 노린 정교한 공격들이 반복적으로 발생할 것으로 보인다. 2026년에는 방치된 '오래된 위험', 즉 EOS된 레거시 시스템에 대한 공격이 크게 증가할 것으로 예상된다. 특히, Windows 10 지원 종료는 이러한 레거시 시스템에 대한 사이버 위협의 기폭제가 될 전망이다.

주요 사고 사례 및 동향

- 예스24 '닷새간 먹통 사태' 원인은?...기술지원 끝난 윈도 서버 OS 썼다(6월)
- 네덜란드, 시트릭스 넷스케일러 취약점 공격으로 심각한 피해 입어(8월)
- 롯데카드, 2017년 공개된 취약점에 당했다...(9월)

네덜란드 국가 사이버 보안 센터는 Citrix NetScaler*의 치명적 취약점(CVE-2025-6543)이 국가의 중요 조직을 침해하는 데 악용됐다고 발표했다. 이는 조직의 외부 경계에 위치한 네트워크 장비의 패치 관리가 미흡할 경우 공격자들의 핵심 진입로가 될 수 있음을 보여줬다. 많은 기업이 서버나 PC 보안에는 집중하지만 정작 '관문' 역할을 하는 네트워크 인프라의 보안 관리가 소홀할 때 어떤 대가를 치르는지 명확히 드러났다.

* 美 SW기업 Citrix의 클라우드 네트워크 플랫폼으로, 네트워크 상에서 어플리케이션 및 데이터를 조정하여 여러 클라우드 서버에 동일한 서비스를 제공하는 로드밸런서 역할 수행



국내에서도 Yes24와 롯데카드 등 대규모 고객 정보를 다루는 기업들의 침해사고가 연이어 발생하며 데이터 유출 위협이 상존함을 다시 한번 확인시켰다. 이러한 사고들은 복잡하게 얽힌 IT 환경 속에서 발생하는 단 하나의 방치된 보안 허점이 기업의 신뢰도와 비즈니스 연속성에 얼마나 큰 타격을 주는지 보여준다.

앞으로도 공격자들은 방어 태세가 집중된 서버 OS 대신 관리가 소홀한 인프라를 우회로로 삼을 것이다. 지원이 종료된 정보자산을 발판 삼아 공격 대상의 내부로 침투 후 측면 이동을 시도할 것이다. 잘 방어된 성문을 공략하기보다는 관리가 소홀한 오래된 성벽을 무너뜨리는 전략이 적극 활용될 것으로 예상된다.

이와 같이 EOS된 레거시 위기를 극복하기 위해서는 첫째, '보이지 않는 자산'을 가시화해야 한다. 내가 무엇을 가졌는지 모르면 무엇을 지켜야 할지 알 수 없다. 조직 내부에 존재하는 모든 IT 자산을 식별하고 특히 Windows 10 운영체제를 포함한 지원 종료(EOS) 대상 시스템의 현황을 정확히 파악하는 것이 대응의 첫걸음이다.

둘째, EOS된 레거시 시스템은 즉시 '격리'해야 한다. 모든 시스템을 즉각 최신화하는 것은 현실적으로 불가능하다. 업그레이드가 불가능한 EOS된 레거시 시스템은 '제로 트러스트' 원칙에 따라 네트워크에서 즉시 분리하거나 최소한의 통신만 허용하는 강력한 마이크로 세그멘테이션(Micro-segmentation)을 적용해야 한다. 이는 하나의 '빈집'이 불타더라도 옆집으로 불이 번지는 것을 막는 방화벽 역할을 한다.

셋째, '패치 관리'의 우선순위를 재정립해야 한다. Citrix NetScaler 사례에서 보듯 서버나 PC뿐만 아니라 네트워크 장비, 하이퍼바이저 등 인프라 전반의 취약점 관리가 중요하다. 특히 외부 접점에 위치하거나 내부망의 핵심 통로가 되는 인프라의 패치 관리는 최우선 순위로 두어야 한다.

3 클라우드 환경의 취약 요소 공격 증가

- 클라우드 전환의 가속화로 공격표면 확장이 보안 위협의 주요 배경으로 작용
- 클라우드-네이티브 공급망 공격, 멀티리전 리스크 등이 새로운 위협으로 부상

클라우드 서비스 이용이 가속화되고, 정보자산의 위치, 상태변화를 파악할 수 있는 가시성이 높아지면서 제어의 복잡성이 급격히 증가하고 있다. 이에 따라 다음과 같이 클라우드 환경에 대한 보안 위협이 발생할 것으로 예상된다.

첫째, AI·자동화 기반 클라우드 침투 공격이다. 2025년 이미 AI 서비스가 활발히 확산되고 있으며, 공격자들은 AI/LLM(대형언어모델)을 활용해 클라우드 환경 스캐닝, 취약점 탐지, 권한 탈취 등의 작업을 자동화할 것으로 보인다. 클라우드-기반 AI 인프라(IaaS, PaaS, SaaS, Docker, Container 등)가 공격자의 새로운 표적이 될 가능성이 높다. 공격 자동화 수준이 올라감에 따라 탐지 및 대응 기술도 고도화시켜야 할 필요가 있다.

둘째, API·컨테이너·서버리스 체인(Chained Exploits) 공격이 증가할 것이다. 개별 취약점이 아닌 API 취약점 → 컨테이너 취약점 → 서버리스 설정오류로 이어지는 연쇄 공격 구조가 클라우드 네이티브 환경에서 현실화할 가능성이 크다. 이는 단일 취약점으로 다수 리소스에 영향을 주는 형태다.

셋째, 클라우드에 특화된 공급망 공격이 확대될 것이다. 특히 CI/CD 파이프라인, IaC(Infrastructure as Code) 템플릿, 서드파티 클라우드 모듈 등이 공격자의 침입 경로가 될 수 있으며, 클라우드 네이티브 환경에 대한 악성 삽입이 증가할 것으로 판단된다.

넷째, 데이터 주권 관련 클라우드 리전(Region) 기반 리스크 확대이다. 멀티 리전/하이브리드 클라우드 운영이 보편화됨에 따라 리전 간 데이터 전송·암호화·접근제어 정책의 불일치가 사고로 이어질 수 있다. 특히 국가별 규제 강화 흐름 속에서 한국 등 특정 지역에서는 리전 기반 보안·거버넌스 준수가 더욱 중요해질 것이다.

다섯째, AI-SOC 시대의 클라우드 로그 조작 및 탐지 회피 위협이다. 클라우드 로그·텔레메트리 데이터를 변조하거나, AI 기반 탐지 모델을 우회하는 공격자가 등장할 수 있으며, 이는 기존 보안 운영센터(SOC)의 탐지 및 대응 역량을 시험할 것이다. 이러한 위협들을 종합하면, 2026년에는 단순한 '설정 오류'나 '권한 남용' 수준을 넘어, 자동화·AI·네이티브 워크로드·공급망·리전 간 거버넌스 등의 복합적이고 전략적인 위협이 클라우드 보안의 주축이 될 것이다.

이러한 클라우드 보안 위협에 대비하기 위해서는

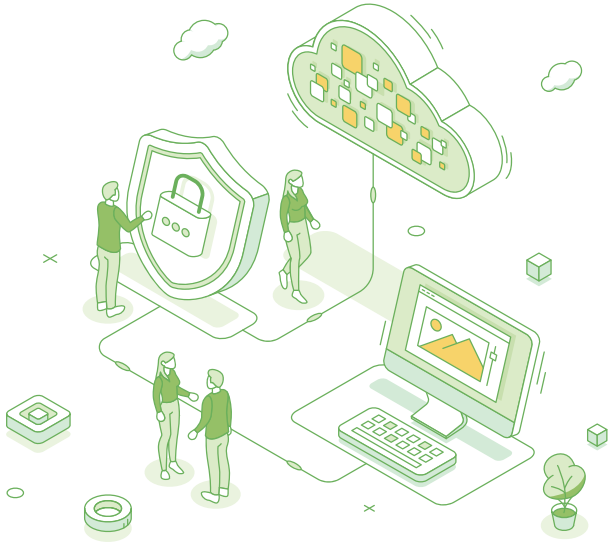
첫째, 리스크 기반 클라우드 보안 운영체계(Risk-Driven Cloud Security)를 정립해야 한다. 자산 식별(ASM) → 위험 점수화 → 우선순위 기반 대응 체계 구축이 필요하다.

둘째, AI-SIEM 및 자율형 보안운영(Autonomous Security Operations)을 도입해야 한다. 클라우드에서 발생하는 방대한 로그와 이상행위를 AI/ML 기술로 실시간 분석하고, 탐지 회피 및 로그 변조 위협에 대응할 수 있도록 해야 한다.

셋째, "Shift-Left" 클라우드 보안 내재화(DevSecOps 강화)이다. IaC 보안, CI/CD 파이프라인 보안 스캔, 컨테이너/서버리스 보안 통합이 필수이며, 개발 단계부터 보안을 내재화해야 한다.

넷째, 클라우드 신뢰모델(Trust Model) 강화 및 거버넌스이다. 데이터 주권, 리전 간 암호화·접근제어, 멀티리전 운영 시 거버넌스 체계 수립이 중요하다.

마지막으로 다섯째, 위협 인텔리전스 통합 및 클라우드-XDR 연계이다. 클라우드 워크로드, API, 컨테이너, 함수(Function) 등 다양한 리소스에 대한 위협 인텔리전스를 실시간 통합하고, 보안 플랫폼(CNAPP, XDR 등)과 연계해 총체적 방어체계를 구축해야 한다.



4 유출된 개인정보를 악용한 2차 사이버 위협

- 해킹, 보안취약점, 내부자 관리부실 등 다양한 요인으로 개인정보 유출 사고 발생
- 개인정보 유출 사태를 악용한 2차 피해 발생 주의 필요

2025년에는 특히 대규모 개인정보 유출 사고도 많이 발생했다. 2월에는 듀오와 GS홈쇼핑이 해킹으로 인해 고객의 이름, 성별, 생년월일, 연락처, 이메일 등을 포함하는 개인정보가 유출됐으며 4월 SKT, 9월 KT 등 이동통신사에서도 악성코드 감염으로 인한 대규모 USIM 정보 유출 사고가 발생했다.

특히 11월에는 쿠팡에서 약 3,370만개에 달하는 대량의 회원정보가 퇴사자를 통해 유출되는 사태가 발생했다.

주요 사고 사례 및 동향

- 결혼정보업체도 터졌다..'듀오' 해킹으로 회원 개인정보 유출(2월)
- GS리테일, 홈쇼핑만 158만건 개인정보 유출(2월)
- 알바몬도 털렸다..."개인 이력서 정보 2만2천건 유출"(5월)
- 쿠팡, 퇴사자 인증키 방치...3370만명 정보 유출 원인 됐다(11월)

과거부터 계속해서 발생해 온 개인정보 유출 사고로 인해 국민들은 '이제 털린만큼 털린게 아닌가', '개인정보는 공공재' 같은 무력감을 표출하고 있다. 그럼에도 불구하고 유출된 개인정보를 악용한 2차 피해 위협에 대해서 면밀한 주의가 필요하다.

다양한 경로에서 수집된 유출 개인정보의 결합은 보이스피싱, 스미싱 등 보다 지능화된 공격에 악용되어 개인정보 유출 피해자들을 추가로 괴롭히고 있다.

공격자들은 '피해보상 신청', '환불' 등이 적힌 안내 문자 또는 해킹으로 인한 '신규카드 발급' 문자 메시지를 위장하여 악성 앱 설치를 유도하거나, '피해사실 조회' 관련 사이트로 위장한 피싱으로 금융정보 탈취를 시도하고 있다. 정보유출 대상자 통보 및 보상 · 환불 절차 안내 등을 빙자한 보이스피싱도 주의가 필요하다.

대규모 개인정보 유출이 발생한 쿠팡 사태와 관련하여 과기정통부는 신속한 조사 및 피해 확산 방지를 위해 민관합동조사단을 가동하였으며, 개인정보위는 쿠팡의 개인정보 보호 의무 위반 여부를 집중 조사하고 있다.
또한, 경찰청 통합대응단에서는 쿠팡을 사칭한 피싱·스미싱 제보를 실시간으로 점검하고 있다. 특히 쿠팡 개인정보 유출 사고 이슈를 악용한 스미싱·보이스피싱 시도 사례*가 지속적으로 발생함에 따라, 한국인터넷진흥원은 보호나라 보안공지를 통해 구체적인 스미싱·보이스피싱 사례를 공개하여 피해가 발생하지 않도록 주의를 당부했다.

* 피해보상 안내 메시지로 위장한 스미싱 메시지에 개인정보가 포함되어 있거나 수사기관 또는 우편물 배송 전화로 위장하여 실제 서비스로 오인할 가능성이 높음

의심되는 URL 또는 홈페이지는 한국인터넷진흥원이 운영하는 카카오톡 채널 '보호나라'의 스미싱·피싱 확인서비스를 통해 간편하게 악성여부 및 신고를 할 수 있으며, 보이스피싱통합신고대응 센터를 통해서도 스미싱 문자 차단 신고할 수 있다. 또한 이동통신사가 무료로 제공하는 '번호도용문자차단서비스'도 사용자 보호에 도움이 될 것이다.

정부는 과기정통부·한국인터넷진흥원·경찰청·금감원 등 관계기관이 긴밀히 협력하여 쿠팡 개인정보 유출 사태로 인한 2차 피해 최소화하기 위해 총력을 기울이고 있으며 개인은 실제 위협 발견시 즉시 경찰이나 한국인터넷진흥원으로 신고하여 피해 발생을 예방하기를 바란다.

맺음말

2025년의 사이버 보안 위협은 기본적인 보안 수칙과 절차가 무시될 때 발생할 수 있는 최악의 사례를 반복적으로 보여줬다. 공격자는 기술과 사람, 공급망과 프로세스를 동시에 노리며, 침해는 한 번의 취약점이 아니라 복합적인 약점의 연결로 발생했다. 특히 소홀한 자산식별 및 내부자 관리, 방치된 시스템 등 과거에는 사소한 결함으로 생각되는 것들이 거대한 방어벽을 와해하게 만드는 결정적 계기가 됐다. 또한 AI의 발전은 사회 모든 분야의 업무 효율화에 기여하는 동시에 공격자들에게는 기존 자신들이 기업 침투를 위한 절차의 반복 행위를 AI로 효율화하여 단시간에 피해를 유발할 수 있는 상황까지 마주하게 됐다.

이처럼 앞으로 공격자들의 AI 기반 자동화된 침해사고 공격 증가와 보다 정교해진 사이버 위협에 효과적으로 대응하기 위해서는 민간과 공공이 영역의 구분 없이 지금보다 더욱 유기적으로 사이버 위협을 탐지 · 공유 · 대응하는 협력체계를 강화해야 한다. 또한 기업들은 랜섬웨어나 공급망 위협 사고 등에 대비하여 협력사와 함께하는 상시 모의침투 훈련이나 보안 교육 등 보다 적극적인 협력체계로 공급망 회복력을 강화하는 것이 더욱 필요할 것으로 분석된다.

마지막으로 기업들은 침해사고가 발생한 경우, 지체없이 과기정통부–KISA에 신고하여 신속하게 사고원인을 분석 · 제거하여 사업을 정상화하고 KISA의 실전형 모의침투 훈련, 보안 취약점 점검, 다양한 예방 서비스를 통해 보안사각지대 최소화 노력 등 보안 방어력을 더욱 높이는 것이 중요할 것으로 생각된다.

하반기

25년 사이버 위협 동향 및 26년 전망

Part 2

Insights / 전문가 칼럼

하반기

25년 사이버 위협 동향 및 26년 전망

Part 2

Insights / 전문가 칼럼

2025년 하반기는 사이버 위협이 기술적 문제를 넘어 사회와 국가 전반의 구조적 리스크로 확산된 시기였다. 대규모 개인정보 유출과 AI 기반 공격의 고도화, 공공 · 민간 핵심 인프라를 둘러싼 연쇄 사고는 기존 보안 체계의 한계를 분명히 드러냈다.

본 보고서의 전문가 칼럼은 이러한 변화의 본질을 진단하고, AI 시대에 요구되는 새로운 보안 패러다임과 대응 방향을 제시하는 데 목적이 있다. AI를 활용한 공격 모델의 진화, 신뢰할 수 있는 AI 거버넌스의 필요성, 침해사고 대응체계와 법 · 제도의 과제, 국가 차원의 정책 변화, 그리고 CISO와 보안조직의 역할 재정의까지 다양한 관점을 담았다.

특히 '신뢰할 수 있는 AI'라는 공통의 화두 아래, AI가 가져온 효율과 위험을 동시에 직시하며, 속도보다 안전과 책임을 우선 하는 보안 체계의 필요성을 강조한다. 이를 통해 현재의 위협 환경을 입체적으로 이해하고, 조직 차원에서 준비해야 할 보안 전략과 정책적 시사점을 도출할 수 있도록 돕고자 한다.

Part. 2

01 2025년 사이버 위협이 바꾼 한국의 AI 보안 지형

'네 개의 충격, 하나의 각성'

이원태 특임교수 – 국민대학교

2025년은 한국 사회가 디지털 기반 위에서 얼마나 불안정하게 서 있는지를 똑똑히 보여준 해였다. 4월 SKT 유심 해킹으로 2,696만 건의 고객정보가 유출됐고, 여름부터 가을까지 KT에서는 불법 펨토셀을 통한 소액결제 사기가 발생했다. 9월 말 국가정보자원관리원(NIRS) 화재로 국가 디지털 행정이 사실상 마비됐으며, 11월에는 쿠팡에서 3,370만 개 계정의 개인정보가 유출되는 사태가 발생했다.

만약 이 사건들이 2010년대에 일어났다면 우리는 아마 "기업 및 기관의 보안관리 부실"이나 "특정 네트워크 장애" 정도로 해석했을 것이다. 그러나 2025년의 사건들은 단순한 해킹 사고가 아니라 사회 시스템적 취약성을 드러낸 구조적 쇼크였다. 이 사건들은 우연히 한 해에 몰려 발생한 것이 아니다. 디지털 시대의 특성과 AI 기반 공격 시대의 도래를 예고하는 사회적 진동이었다.

어쩌면 2025년은 우리 사회가 비로소 "AI 사회에 걸맞은 AI 보안 체계가 필요하다"는 명제를 뼈저리게 체득한 시기였다고 할 수 있다. 우리는 지난 10여 년 동안 디지털 전환을 주저없이 추진해 왔다. 전자정부, 전자상거래, 디지털 금융, 모바일 신원인증까지 한국은 세계에서 가장 빠른 속도로 디지털 사회로 이동해 온 나라였다. 하지만 2025년의 사건들은 이 빠른 이동 속에 이미 깊숙이 스며든 구조적 취약성이 존재했다는 사실을 드러냈다는 것이다.

신뢰가 무너진 네 계절, 2025년의 기록

신뢰의 붕괴는 봄과 함께 찾아왔다. 4월 SKT에서 발생한 2,696만 건의 유심(USIM) 정보 해킹 사건은 우리 사회가 믿고 있던 연결의 안전판이 얼마나 취약한지 보여준 첫 번째 충격이었다. 유심은 단순한 칩이 아니라 스마트폰 보안의 최후의 보루이자 디지털 신분증이다. 국제이동통신가입자식별번호(IMSI)를 포함한 핵심 인증 정보의 탈취는 단순한 개인정보 유출을 넘어, SIM 스와핑(SIM Swapping)이나 통화 가로채기 같은 2차 공격의 길을 열어줬다. 공격에 사용된 악성코드는 리눅스 서버를 노린 고도화된 백도어로, 네트워크 방화벽을 우회할 수 있는 기능을 가지고 있어 장기간 잠복하면서 정보를 유출할 수 있었다.

여름에는 통신망 설계 자체의 허점이 드러났다. 8월부터 경기 광명, 서울 금천 등 수도권 지역에서 KT 이용자들의 무단 소액결제 피해 신고가 이어졌다. 범인들은 불법 펨토셀(초소형 무선 기지국)을 차량에 싣고 새벽 시간대에 대규모 아파트 단지를 돌아다니며 인근 주민들의 휴대폰을 강제로 불법 기지국에 연결시켜 SMS 인증 문자를 가로채고 소액결제를 진행했다. 총 20개의 불법 펨토셀이 확인됐고, 약 2만 2천여 명이 이들 기지국에 접속했다. 특히 주목할 점은 일부 펨토셀이 300일 이상 활동했으나 소액결제 피해는 없었다는 사실이다. 이는 범죄조직이 단순 소액결제 사기보다 더 큰 목적—예컨대 대규모 통신 데이터 수집—을 노렸을 가능성을 시사한다. 펨토셀 인증서가 유효기간 10년으로 설계돼 있었고, 네트워크 접속 구조가 단순해 비인가 기지국도 망에 연결될 수 있었다는 점은 통신인프라의 보안이 물리적 장비를 넘어 인증 체계와 네트워크 설계 전반에 걸쳐 재검토돼야 함을 보여줬다.

가을에는 디지털 국가 행정이 멈췄다. 9월 26일 대전 국가정보자원관리원(NIRS)에서 발생한 리튬배터리 화재는 사이버 공격이 아니었음에도 디지털 정부 전체를 마비시켰다. 배터리 교체 작업 중 발생한 물리적 화재 하나로 정부24와 모바일 신분증 등 수백 개의 공공서비스가 먹통이 됐다. 대전 본원에 전체 국가 정보시스템의 3분의 1 이상이 집중돼 있었고, 재해복구(DR) 체계는 실질적으로 작동하지 않았다. 일부는 저장소만 있거나 백업만 있는 경우도 있었다. 이 사건은 AI 행정망을 논하기 전에 불이 나면 꺼지는 서버부터 걱정해야 하는 현실을 적나라하게 드러냈다.

혼란의 정점은 겨울 초입인 11월에 찍혔다. 쿠팡에서 3,370만 명의 개인정보가 유출됐다는 사실이 밝혀진 것이다. 성인 인구 대부분을 포괄하는 규모였지만, 더 무서운 것은 그 내용이다. 구매 이력, 배송지, 연락처 등은 생성형 AI가 피싱 공격이나 사기 시나리오를 작성할 때 가장 탐내는 기초 데이터이다. 사건의 원인은 퇴사한 중국인 직원의 액세스 토큰 서명키가 방치된 것으로 퇴사자의 접근 권한을 제때 회수하지 않은 내부 통제 실패가, 역설적으로 전 국민을 대상으로 한 AI 범죄의 토양을 제공한 셈이 됐다.

이처럼 2025년은 단순히 '피해의 해'가 아니라 '드러남의 해'였다고 해도 과언이 아니다. 사건들의 파괴력보다 중요한 것은 그 사건들이 무엇을 드러냈는가이다. 통신 기반 신원체계의 취약성, 망 설계의 구조적 허점, 공공 인프라 복원력의 부재, 내부 거버넌스 실패, 데이터 통제 능력 상실, 그리고 AI 공격 시대의 위험성까지, 이 모든 것이 "2025년 하나의 연속 선상"에서 일어났다는 사실은 우연이 아니다. 그런 점에서 2025년은 한국 사회가 AI 시대에 필요한 보안·인프라·정책 체계를 갖추지 못한 상태에서 초연결 디지털 사회로 진입했음을 보여준 해였다고 할 수 있다.

AI 공격이 아니었기에 더 무서운 이유

이들 사고는 AI 공격과 직접적 연관은 없었지만, 그렇기에 더욱 위협적이다. 지금 수준의 공격도 이 정도라면, AI가 공격자 편에 서는 순간 어떤 결과가 벌어질지 쉽게 상상할 수 있다. 그 양상과 충격은 AI 공격 시대의 위험성을 미리 보여주는 전조이다. 사건 하나하나는 서로 다른 원인을 갖고 있었지만, 이들이 만들어내는 곡선은 똑같은 방향을 가리키고 있었다. 한국 사회는 단일하고 지나치게 넓은 디지털 위험표면 위에 서 있다는 것이다.

사고들의 공통점은 명확했다. 단일 지점에서 사고가 발생해도 사회 전체 기능이 흔들렸다는 것이다. SKT 유심 해킹은 통신 기반 인증 전반을 위협했고, KT 펨토셀 사건은 소액결제와 휴대폰 인증, 위치기반 서비스를 교란시켰다. NIRS 화재는 중앙정부의 수백 개 디지털 서비스를 마비시켰고, 쿠팡 유출은 소비자 데이터 전반의 신뢰를 붕괴시켰다. 2025년 이전에는 한 기관의 사고는 그 기관의 문제였다. 하지만 올해는 한 기관의 사고가 사회 전체의 문제가 됐다.

더 중요한 것은 이 사건들이 AI 공격자가 가장 선호할 정보와 구조를 노출시켰다는 점이다. SKT는 신원과 인증 기반 데이터를, 쿠팡은 타깃팅 기반 데이터를, KT는 SMS 인증 구조와 망 구성 정보를, NIRS는 공공 인프라의 물리적·논리적 구조 자체를 노출했다. 이 네 사건이 조합되면 AI 공격이 인간 공격보다 수십 배 정교한 공격 시나리오를 설계할 수 있다.

많은 사람들이 "왜 사건들을 AI 공격으로 몰아가느냐"고 반문할 수 있다. 그러나 전문가의 관점에서 2025년에 발생한 사건들이 가장 무서운 이유는 AI 공격이 아니었기 때문이다. 만약 이 사고들에 취약점 스캐닝 자동화, 공격 전술 자동 조합, 사회공학 메시지 맞춤형 생성, 대규모 병렬 공격 수행이 더해졌다면 파괴력은 지금의 5배에서 10배는 되었을 것이다. 즉, AI 공격 시대에 발생할 일의 초기 버전이었고, 시뮬레이션 또는 튜토리얼과 같은 성격을 띤 것이라 할 수 있다.

마침 2025년 11월 미국 Anthropic의 발표는 이러한 우려가 기우가 아님을 보여줬다. Anthropic은 자사 모델 Claude와 Claude Code가 중국계 공격조직의 실제 사이버 공격 과정에 사용됐다는 사실을 공개했다. AI는 자동 취약점 스캐닝, 환경 기반 공격 코드 생성, 내부 이동 전략 설계, 피해자 맞춤형 피싱 메시지 생성, 단계별 공격 시퀀스 제안을 수행했다. 이 발표는 사이버 위협의 본질이 근본적으로 달라진다는 것을 보여준다. AI의 폭발적 확산과 함께 AI를 악용한 공격의 자율화(Autonomous Offense)가 본격화될 것이란 예고인 셈이다. 한국 사회는 아직 인간 공격자에게도 취약한데, AI 공격자가 등장하면 훨씬 더 취약해질 수밖에 없다. 그런 점에서 2025년의 사건들은 그 현실을 예고한 전조적 징후였다.

국가 AI 거버넌스의 대전환 : 정책과 제도의 재편

2025년의 사건들은 AI 시대를 준비하는 보안 체계가 얼마나 시급한지를 명확히 보여줬다. 다행히 정부도 이 위기 신호를 감지하고 기민하게 움직이기 시작했다. 2025년의 사건들이 한국 디지털 사회의 취약성을 드러냈다면, 같은 해 정부가 추진한 정책들은 그 취약성이 단순한 IT 이슈가 아니라 국가 운영 방식과 공공 인프라의 구조 전체를 다시 설계해야 하는 문제임을 인식하기 시작한 것이다. 많은 사람들은 "대형 사고가 많았던 해"라고 말하지만, 정책 분석의 관점에서 2025년은 한국 디지털 거버넌스 체계가 AI 중심으로 재편되기 시작한 해라고 보아야 할 것이다. 정책의 초점이 기술 확산에서 기술 안전성으로, AI 도입에서 AI 위험관리로, 정보화에서 복원력 중심 구조로 전환하기 위한 모색의 시기였다. 이는 우연히 등장한 단발 정책들의 집합이 아니라, 명확한 방향성을 갖춘 구조적 전환의 시작이었다.

2025년 정부조직 개편에서 가장 눈에 띄는 변화는 대통령실에 AI수석비서관실이 신설된 것이다. 이는 AI를 단순히 잘 활용하겠다는 의미가 아니라 AI를 국가전략의 핵심 의제로 격상시키는 선언이다. AI는 더 이상 산업정책의 한 계열이 아니다. 통신, 금융, 국토, 복지, 교육, 문화, 보건 어떤 분야든 AI가 정책효과와 행정방식에 근본적 영향을 준다. 따라서 AI는 과기정통부 단일 부처의 범위를 넘어 국가 차원의 정책 조정과 통합 전략이 필요한 영역으로 발전했다. 이같이 AI 정책의 범정부적 중요성이 강조되면서 과기정통부 역할이 부총리로 격상된 것도 이러한 흐름의 일환이다.

AI수석의 등장만큼 중요한 것은 각 부처별로 Chief AI Officer(CAIO)를 지정하고 범부처 CAIO 협의체를 구성한 점이다. 이는 미국 연방정부가 이미 수년 전부터 시행하고 있는 체계이며, 한국도 이를 본격적으로 제도화한 것이다. CAIO 협의체는 AI가 더 이상 기술이 아니라 행정 운영 방식 자체가 됐음을 의미한다. 각 부처의 실행 프로세스가 AI 기반 정책을 중심으로 재편되고, AI 위험관리 체계가 부처 단위가 아니라 정부 전체 단위가 되는 것이다.

2025년 9월 8일에 출범한 국가AI전략위원회는 한국 AI 정책사에서 가장 중요한 거버넌스 설계라고 평가받는다. 이 위원회는 단순한 정책 자문기구가 아니라 산업, 규제, 안전, 거버넌스, 윤리, 공공서비스, 국제협력을 모두 아우르는 AI 국가전략의 통합 조정 기능을 맡는다. 앞으로 국가AI전략위원회는 단순히 AI 기술의 도입을 넘어 AI 기반 사회 운영 방식을 설계할 것이며, 그런 점에서 조만간 발표 예정인 국가 AI 액션플랜은 단순한 산업계 청사진이 아니라, AI 보안 정책 과제들까지 포함하는 종합 전략이 될 것으로 예상된다. 왜냐하면 AI 보안은 AI 산업보다 더 기본적인 "AI 사회의 지속가능성 조건"이기 때문이다. 그런 점에서 AI 액션플랜은 단순한 산업계 청사진이 아니라 AI기본사회이자 AI 안보사회의 설계도가 될 것이다.

또한 2025년 정부 정책 중 가장 깊은 의미를 가진 것은 9월 30일 국정원에서 발표한 국가망보안체계(N²SF)다. N²SF는 단순한 기술 가이드라인이 아니다. 한국의 보안사에서 2000년대식 망분리 체계를 19년 만에 폐기한 사건이다. 기존 망분리는 보안을 위해 외부 인터넷 단절을 전제로 했지만, AI, 클라우드, 원격근무, 데이터 기반 업무환경에서는 더 이상 지속할 수 있는 모델이 아니었다. N²SF가 제시한 방향은 명확하다. 보안의 단위를 망에서 데이터로 이동하고, 망 차단이 아니라 데이터 보호, 접근통제, 권한관리 중심으로 재설계한다. 업무 중요도와 기밀 수준에 따라 기밀, 민감, 공개로 구분하여 보안 등급을 차등 적용하며, 크로스도메인솔루션(CDS) 기반의 영역 간 통제를 구현한다. N²SF는 공공부문의 업무방식 전체를 바꾸게 될 것이다.

뿐만 아니라 2025년 10월 22일에 발표된 정보보호 종합대책은 연쇄 사고들에 대한 정부의 가장 중요한 정책적 응답이다. SKT, KT, NIRS 등 연이은 보안 사고에 대응하여, 정부는 국가안보실을 중심으로 과기정통부, 금융위원회, 개인정보보호위원회, 국정원, 행정안전부 등이 합동으로 대책을 발표한 것이다. 일단 정부는 선제적 조치의 일환으로 공공 기반시설 288곳, 중앙·지방기관 152곳, 금융업 261개, ISMS 기업 949개 등 총 1,600개 핵심 IT 시스템에 대한 전수 점검을 실시했다. 유출 위험이 높은 데이터부터 우선적으로 보호하는 데이터 중심 보안으로 전환하고, AI 레드팀 구성과 AI 기반 탐지·대응 체계를 구축하겠다고 밝혔다.

특히 AI 공급망 보안체계의 구축은 향후 AI 사회의 신뢰를 지탱하는 근본적인 인프라가 되며, 2026년 이후 한국 디지털 정책의 핵심 축이 될 것이다. AI 시대의 보안은 더 이상 모델 자체의 안전성만으로 설명되지 않는다. AI 시스템은 모델, 데이터셋, 외부 API, 오픈소스 컴포넌트, 학습 인프라, 업데이트 체계 등 수많은 요소가 결합된 다층적 공급망(Supply Chain) 위에서 작동한다. 2025년 KT 펨토셀 사기, SKT 유심 해킹, 쿠팡 3,370만 개인정보 유출 사례가 보여줬듯이 단 하나의 취약점—방치된 인증키, 부정사용된 기지국, 검증되지 않은 오픈소스—이 전체 시스템의 신뢰를 무너뜨리고 사회 전체의 위험으로 확장될 수 있다. 이러한 구조적 취약성은 AI 시대에 더욱 심화되며, 특히 AI 모델이 외부 오픈소스·데이터·API에 의존하는 만큼 AI 공급망 보안은 산업경쟁력·국가안보·수출규제 대응의 핵심 인프라가 된다. 따라서 향후 정부는 모든 AI·SW 구성요소를 투명하게 기록·관리하는 SBOM 기반 AI 공급망 보안 표준을 국가 차원에서 제도화하고, 의료·교통·금융 등 분야별 AI 공급망 보안 프레임워크를 마련해 산업별 요구에 맞는 정교한 가이드라인을 제공할 것으로 보인다.

또한 2025년 12월 10일, 국정원과 과기정통부는 동일한 날 서로 다른 타깃을 대상으로 공공·민간 AI 보안 가이드라인을 동시에 발표했는데, 이는 AI 시대의 위험을 수용할 수 있는 새로운 보안·안전성 체계를 설계해야 한다는 국가적 과제임을 인식하기 시작한 신호라고 볼 수 있다. 특히 주목해야 할 점은, 기존의 네트워크·물리 인프라 중심 규범을 넘어 AI 시스템 자체의 위협·공급망·생애주기 전체를 아우르는 새로운 보안 기준이 등장했다는 점이다.

먼저 국가인공지능안보센터는 공공기관을 대상으로 생성형·에이전틱·피지컬 AI 위협을 포함한 15개 보안위협과 30개 보안대책, 그리고 AI 생애주기별 보안 체크리스트 57개를 제시하는 실무형 가이드북을 발표했다. 이는 기존의 "망 중심 보안"으로는 대응할 수 없는 AI 위협의 특성을 반영하면서, 공공부문에서 AI 시스템 안전 도입을 위한 최초의 구조적 기준을 제공했다는 점에서 한국 보안정책사에 중요한 전환점이라 하겠다. 특히 이 가이드북은 AI 모델, 학습데이터, 에이전트 권한, 외부 API 연동, 공급망 보안 등 그동안 공공 규제에서 다루기 어려웠던 핵심 위협을 공식 항목으로 포함했고, 공공기관이 실제로 운영 가능한 형태로 설계되어 AI 시스템 도입→개발→운영→폐기 전 과정에 대한 최초의 공공보안 레퍼런스로 평가할 수 있다.

또한 과기정통부와 KISA가 같은 날 발표한 'AI 보안 안내서'는 기업·개발자·서비스 제공자·이용자 모두가 참고할 수 있는 민간용 보안 기준으로, 총 113개의 보안 요구사항을 통해 AI 모델 개발 단계의 Secure-by-Design, 데이터 암호화, 공급망 관리, 실시간 이상탐지, API 보안, 서비스 복원력 확보 등을 포괄했다. 이는 한국에서 처음으로 "AI 보안"을 전통적인 정보보호가 아닌 독립된 기술·운영 영역으로 구분하고, AI 모델·데이터·API·Inference Layer·Agent Layer 등을 통합적인 사이버 시스템으로 간주해 위험을 정의한 최초의 범정부 보안 기준이라 할 수 있다.

두 문서는 성격은 다르지만, 공공·민간 모두에 대해 AI 보안의 '국가 기준'을 제공했다는 점에서 하나의 큰 정책 흐름을 형성한 것이라 하겠다. 즉, AI 보안이 더 이상 자율적 모범사례 수준이 아니라 국가적 안전성 확보의 규범 영역으로 편입되었다는 의미이다. 특히 AI 공격 및 AI 공급망 리스크가 기존 사이버 공격보다 5~10배 더 빠르고 정교해질 수 있다는 전망 속에서, 정부가 공공과 민간을 구분해 이중 구조의 AI 보안 체계(국정원 공공용 기준 + 과기정통부 민간용 기준)를 정립하기 시작했다는 점은 향후 2026년 국가 AI 안전성 기준 및 AI 보안 표준화 체계의 기초가 될 것으로 보인다.

이상과 같이 정부의 AI 정책 및 제도 변화는 단순한 사고 대응이 아니라, AI 시대의 국가보안·데이터보안·AI 공급망 보안을 전면 재설계하는 국면 전환의 출발점이라고 볼 수 있다. AI수석 신설, CAIO 협의체, 국가AI전략위원회 출범, N²SF, 정보보호 종합대책이 AI 보안 거버넌스의 기반을 세웠다면, 늦은 감 있지만 최근에 발표된 국정원의 '공공 AI 보안 가이드북'과 과기정통부·KISA의 'AI 보안 안내서'는 AI 사회의 안전성을 위한 기술적·운영적 규범의 첫 공식 버전이라고 할 수 있다. 특히 이 두 문서는 AI 시스템 생애주기(Lifecycle) 기반 위험관리, AI 공급망(Supply Chain) 보안, 데이터 중심 보안 모델, 에이전트·모델·API 권한관리 강화, AI 사고 대응·모니터링 체계 구축을 국가적 규범으로 도입함으로써, 2026년 새롭게 정립할 국가 차원의 AI보안 프레임워크의 서장을 연 것으로 볼 수도 있다. 이는 2025년의 위기가 단순한 해킹 증가가 아니라 AI 기본사회에서의 새로운 보안 철학, 즉 '운영·책임·복원력 중심의 AI 보안'이 국가적 규범으로 자리 잡기 시작했다는 뜻이기도 하다.

유관기관 및 산업계의 대응

2025년의 연이은 대형 사고는 기술적 취약점에 앞서 우리 사회 전반에 깊이 스며든 '보안 불감증'이 더 근본적인 문제임을 드러냈다. 대기업과 중소기업을 막론하고 6,000만 건이 넘는 개인정보가 유출된 것은 단일 기업의 실패가 아니라, 2000년대 초반에 설계된 한국의 보안 패러다임이 초연결·AI 기반 환경에 적응하지 못한 채 유지되어 왔기 때문이다. 전문가들은 올해 사고들이 모두 기본적인 통제—퇴사자 권한 회수, 인증키 관리, 최소권한 원칙, 설계 검증—에서 무너졌다는 점을 공통적으로 지적한다. 이러한 배경에서 제로트러스트(Zero Trust)의 필요성은 더욱 뚜렷해졌다. 제로트러스트는 내부와 외부를 가리지 않고 모든 접근을 의심하고 검증하는 현대적 보안 패러다임으로, 2025년 발생한 주요 사고 상당수는 이 원칙이 제대로 적용됐다면 예방 가능했을 사건들이었다. 그러나 국내에서는 여전히 보안을 비용이나 불편으로 인식하고, 보안 인력 처우 역시 낮아 우수 인재가 유입되지 못하는 구조적 한계가 지속되고 있다. 따라서 정책적 접근은 단순한 규제나 사후 처벌을 넘어, 기본적 의무를 강제하면서도 보안을 강화한 기업에 인센티브를 제공하는 '당근과 채찍' 체계로 전환돼야 한다는 지적이 힘을 얻고 있다. 결국 보안문화의 재정립 없이는 기술적 보안 강화나 거버넌스 개편도 한계를 가질 수밖에 없다.

또한 정보보호 종합대책에서는 CEO와 CISO의 법적 책임과 권한을 강화해 전사 보안 정책의 승인권, 예산통제권, 이사회 보고 의무 등을 명문화한 점이 눈에 띈다. 이는 보안을 기술부서의 문제가 아닌 '경영 의사결정의 핵심

축'으로 재위치시키는 의미로, 단순 사고 예방을 넘어 기업의 지배구조 자체가 보안 중심으로 재정렬되는 흐름을 보여준다. 특히 AI 기반 서비스가 경영 전반에 스며드는 상황에서, 최고경영진이 보안 리스크를 직접 책임지는 구조를 제도화했다는 점은 향후 기업 신뢰성과 지속가능성의 기준이 크게 변화하고 있음을 시사한다.

그외에도 기업의 보안 수준을 공식적으로 등급화해 공개하는 정보보호 등급제를 도입하고, 공공부문은 정보화 예산 대비 일정비율 이상을 보안에 투자하도록 의무화하며, 보안 최고 전문가인 화이트해커를 연간 500여 명씩 양성하되, 기업 수요에 맞춰 양성 체계를 재설계하는 등 여러 정책과제들도 제시했다. 이번 종합대책은 현 사안의 시급성을 고려하여 즉시 실행할 수 있는 단기 과제 위주로 제시됐으며, 중장기 과제를 망라하는 국가 사이버안보 전략을 연내 수립할 계획이라고 밝혔다.

물론 기업과 기관이 준수해야 할 요건도 명확해졌다. 정보통신망법과 개인정보보호법에 따른 ISMS 인증 의무 기업은 정보보호 공시를 의무화하며, 상장사 전체로 확대 적용된다. CISO와 CPO는 독립적 권한을 보장받되 동시에 더 큰 책임을 지게 되었다. 정보보호 예산은 전체 IT 예산의 일정 비율 이상을 확보해야 하며, 이는 감사 항목에 포함된다. AI 시스템을 도입하는 기업은 사전 위험평가를 수행해야 하고, 고위험 또는 고영향 AI로 분류될 경우 추가적인 안전성 검증과 모니터링 체계를 갖추어야 한다. 내부 통제 실패로 인한 개인정보 유출 시 과징금이 대폭 상향되었고, 경영진의 직접적 책임도 강화됐다.

NIRS 화재 이후 정부와 공공기관은 단순히 보안 체계를 강화하는 것 이상의 과제, 즉 국가 디지털 복원력 체계 재구축의 필요성을 절감했다. 완벽한 방어는 존재하지 않으며, 사이버 공격뿐 아니라 물리적 재난도 공공 디지털 서비스를 무너뜨릴 수 있고, AI 기반 공격은 탐지를 불가능하게 만들 수 있기 때문이다. 정부가 직면한 핵심 과제는 민간 클라우드 활용 확대, 다중화된 공공 인프라 구축, 재해복구와 업무연속성의 실질적 작동, 공공데이터의 분산 아키텍처, AI 기반 자동 복구 시스템 도입 등이다.

2025년의 정책 전환은 유관기관의 역할과 기능을 근본적으로 재정의하는 계기가 됐다. 특히 한국인터넷진흥원(KISA)과 한국지능정보사회진흥원(NIA)은 AI 시대의 보안과 AI 전환을 실질적으로 지원하는 핵심 기관으로 자리매김할 것으로 기대된다.

KISA는 2025년 하반기부터 AI 보안 전담조직을 신설하고 AI 기반 침해사고 대응 체계를 본격화했다. 기존의 사이버 위협 인텔리전스 시스템에 AI 기반 이상탐지 기능을 통합하고, 대규모 언어모델(LLM)을 활용한 피싱 메시지 자동 분석 시스템을 구축했다. SKT와 KT 사건 이후에는 통신사 보안 감독 기능을 강화해 분기별 통신 인프라 보안 점검을 의무화하고, 펨토셀과 같은 소형 기지국에 대한 인증 체계 개선 작업을 주도했다. 또한 AI 시스템 취약점 신고 포상제를 신설해 AI 모델과 서비스에서 발견된 보안 취약점을 민간 보안 연구자들이 신고할 수 있는 창구를 마련했다.

특히 주목할 만한 것은 KISA가 운영하는 사이버 위기경보 체계에 'AI 위협' 항목을 신설한 것이다. 기존의 관심–주의–경계–심각 4단계 경보 체계에 AI 기반 공격 징후를 별도로 모니터링하는 시스템을 추가했다. 이는 Anthropic 사례처럼 AI가 실제 공격에 사용되는 상황을 조기에 탐지하고 대응하기 위한 조치라고 할 수 있다.

또한 NIA는 공공부문의 AI 도입과 디지털 전환을 지원하는 기관으로서 2025년 NIRS 화재 이후 공공 인프라 복원력 강화 사업을 주도하고 있다. 중앙부처와 지방자치단체를 대상으로 재해복구(DR) 체계 실태조사를 실시하고, 업무연속성계획(BCP) 수립 가이드라인을 배포했다. 특히 클라우드 기반 재해복구 서비스 도입을 지원하기 위해 민간 클라우드 사업자와의 협력 체계를 구축하고, 공공기관의 클라우드 전환 시범사업을 확대했다.

또한 공공 AI 서비스 품질관리 체계를 구축하는 역할을 맡고 있다. 정부 부처가 도입하는 AI 시스템에 대한 사전 영향평가를 지원하고, AI 모델의 편향성과 공정성을 검증하는 도구를 개발·보급하고 있다. 2025년 하반기부터는 공공 AI 서비스 레지스트리를 구축해 어떤 부처에서 어떤 AI 시스템을 사용하고 있는지 통합 관리하기 시작했다. 이는 향후 AI 사고 발생 시 신속한 대응과 영향 범위 파악을 위한 기초 인프라가 될 것이다.

산업계의 변화도 뚜렷하게 나타나기 시작했다. 금융권에서는 2025년 하반기부터 AI 거버넌스 체계 구축이 본격화됐다. 주요 은행들은 'AI 윤리위원회'를 신설하고, AI 모델 개발부터 배포, 운영, 폐기까지 전 생애주기를 관리하는 체계를 도입했다. KB금융과 신한금융은 AI 리스크 관리 전담조직을 신설했고, 외부 AI 모델 도입 시 공급망 보안 검증을 의무화했다. 금융감독원은 2026년부터 AI 활용 금융서비스에 대한 별도 감독 기준을 마련할 예정이며, 이에 대비한 금융권의 준비가 한창이다.

제조업계에서는 스마트팩토리와 AI 기반 생산관리 시스템의 보안이 화두가 됐다. 삼성전자와 현대자동차는 2025년 하반기 AI 보안 가이드라인을 사내 표준으로 채택하고, 협력사에도 동일한 기준을 적용하기 시작했다. 특히 공급망 전체의 AI 시스템 보안을 관리하기 위해 협력사 대상 AI 보안 교육 프로그램을 운영하고, 정기적인 보안 점검을 실시하고 있다.

통신업계는 2025년 사건들의 직접적 당사자로서 가장 큰 변화를 겪었다. SKT와 KT는 CISO 직급을 부사장급으로 상향하고, 정보보호 예산을 전년 대비 50% 이상 증액했다. 통신 3사는 공동으로 'AI 시대 통신보안 협의체'를 구성하여 AI 기반 위협 정보를 실시간으로 공유하고, 공동 대응 체계를 마련했다. 펨토셀 사건 이후에는 소형 기지국에 대한 인증 체계를 전면 개편하고, 네트워크 접속 시 다단계 인증을 도입했다.

플랫폼 기업들도 내부 통제 체계를 대폭 강화했다. 쿠팡 사건 이후 네이버, 카카오, 배달의민족 등 주요 플랫폼 기업들은 임직원 권한 관리 시스템을 전면 개편했다. 퇴사자의 접근 권한을 즉시 회수하는 자동화 시스템을 도입하고, 핵심 데이터 접근 시 다중 인증과 승인 절차를 의무화했다. 또한 AI 기반 이상행위 탐지 시스템을 도입하여 비정상적인 데이터 접근이나 대량 다운로드 시도를 실시간으로 모니터링하고 있다.

중소기업을 위한 지원도 확대되었다. KISA와 NIA는 중소기업 대상 AI 보안 컨설팅 프로그램을 운영하고, AI 시스템 도입 시 필요한 보안 체크리스트와 가이드를 무상으로 제공하고 있다. 정부는 2026년부터 중소기업의 AI 보안 투자에 대한 세액공제를 확대하고, ISMS-P 인증 취득 비용을 지원하는 사업을 확대할 예정이다.

CISO 및 보안담당자가 알아야 할 것들

2025년의 연이은 사건들은 조직 내에서 보안이라는 기능 자체가 어떤 역할을 맡아야 하는지를 다시 질문하게 만들었다. 특히 기업의 최고정보보호책임자(CISO), 공공기관의 정보보호 담당자, 그리고 점점 증가하고 있는 AI 안전성, AI 책임성 담당 부서의 리더들에게 2025년의 사건들은 하나의 냉혹한 메시지를 던졌다. 기존의 정보보안 패러다임은 AI 시대에 충분하지 않다는 것이다.

정보보안의 역사를 간단히 되짚어 보면, CISO의 역할은 시대에 따라 크게 세 번 진화해 왔다. 2000년대 초기에는 침입 방지와 망분리가 중심이었고, 2010년대에는 개인정보보호와 클라우드 전환에 대응했으며, 2020년대 후반에는 AI 기반 서비스와 초연결 인프라의 보안을 총괄하게 되었다. 그러나 이제 다가오는 시대는 완전히 다른 국면을 요구한다. CISO는 더 이상 취약점 점검 보고서를 작성하거나 로그 데이터를 모니터링하거나 보안 솔루션을 구매하고 배포하는 역할이 아니다.

AI 시대의 CISO는 조직 내에서 위험을 분석하고, 위험을 번역하고, 위험을 구조화하며, 궁극적으로 위험을 통치하는 책임자가 되어야 한다. CISO는 조직의 정보보안 담당자가 아니라 조직의 신뢰를 책임지는 관리자가 된다. AI 시대의 위험은 기술적 위험이 아니라 구조적 위험이기 때문이다. 2025년 사고의 본질은 개별 기술 실패가 아니라, 그보다 훨씬 근본적인 구조적 취약성이었다.

AI 보안 시대의 위험은 여러 층위로 나타난다. 데이터 출처가 불분명하거나 불균형과 편향이 있고 개인정보가 혼입되거나 변조될 수 있는 데이터 위험이 있다. 환각이나 과적합, 과신, 미검증 모델 재사용, 모델 중독 같은 모델 위험이 있으며, AI 에이전트가 권한을 초과한 행동을 수행하거나 외부 API 연결을 통해 예측 불가능한 행동을 할 수 있는 에이전트 위험도 있다. 외부 모델과 데이터셋에 의존하고 오픈소스 모델의 취약성이나 서드파티 API 리스크가 존재하는 공급망 위험, 그리고 내부 통제 실패와 역할·책임 미정의, 부서 간 분절적 의사결정 같은 조직 거버넌스 위험까지 복합적으로 얽혀 있다.

보안담당자와 CISO가 당장 준비해야 할 구체적인 사항들을 정리하면 다음과 같다.

첫째, AI 시스템 인벤토리를 구축해야 한다. 조직 내에서 사용 중인 모든 AI 시스템, 모델, 서비스를 파악하고 문서화해야 한다. 어떤 부서에서 어떤 AI를 사용하는지, 그 AI가 어떤 데이터에 접근하는지, 외부 API와 연결되어 있는지 등을 명확히 파악하지 못하면 위험관리 자체가 불가능하다.

둘째, AI 공급망 관리 체계를 수립해야 한다. 외부에서 도입한 AI 모델의 출처와 학습 데이터, 알려진 취약점, 라이선스 조건 등을 검증하는 프로세스가 필요하다. 오픈소스 AI 모델을 사용할 경우 해당 모델의 보안 패치 이력을 지속적으로 모니터링해야 하며, 상용 AI 서비스를 이용할 경우 공급업체의 보안 수준과 SLA를 명확히 확인해야 한다.

셋째, 데이터 분류 및 접근통제 체계를 재정비해야 한다. N²SF가 제시한 것처럼 데이터를 기밀, 민감, 공개로 분류하고, AI 시스템이 각 등급의 데이터에 접근할 때 어떤 통제가 필요한지 정의해야 한다. 특히 AI 학습에 사용되는 데이터에 개인정보나 민감정보가 포함되지 않도록 데이터 익명화와 가명처리 프로세스를 강화해야 한다.

넷째, AI 사고 대응 체계를 구축해야 한다. 기존의 사이버 침해사고 대응 체계와는 별도로 AI 특화 사고 대응 절차가 필요하다. AI 모델이 비정상적인 출력을 생성하거나 편향된 결과를 내놓을 때, AI 에이전트가 예기치 않은 행동을 할 때, AI 시스템이 해킹당했을 때 각각 어떻게 대응할지 사전에 정의하고 훈련해야 한다.

다섯째, 임직원 권한 관리를 자동화해야 한다. 쿠팡 사건은 퇴사자 권한을 회수하지 않은 내부 통제 실패가 핵심 원인이었다. 모든 시스템 접근 권한은 재직 상태와 자동으로 연동돼야 하며, 퇴사 처리 즉시 모든 접근 권한이 회수돼야 한다. 특히 관리자 권한이나 데이터베이스 직접 접근 권한 같은 고위험 권한은 정기적으로 재검증하고, 사용 이력을 감사해야 한다.

여섯째, AI 윤리·안전 검토 프로세스를 제도화해야 한다. 새로운 AI 시스템을 도입하거나 기존 AI 시스템을 고도화할 때 보안 검토뿐 아니라 윤리적·사회적 영향 평가를 수행해야 한다. 이를 위해 법무, 인사, 준법감시, 정보보호 부서가 참여하는 AI 거버넌스 위원회를 구성하는 것이 바람직하다.

일곱째, CISO의 권한과 책임을 명확히 해야 한다. CISO는 단순 보고 라인이 아니라 실질적인 의사결정 권한을 가져야 한다. AI 시스템 도입에 대한 보안 승인권, 정보보호 예산에 대한 직접 통제권, 이사회 직접 보고 채널 등이 보장되어야 하며, 동시에 보안 사고 발생 시 CISO의 책임 범위도 명확히 정의돼야 한다.

실무적으로 2026년을 준비하는 보안담당자가 체크해야 할 리스트는 다음과 같다. 정보보호 공시 대상 기업인 경우 공시 항목을 점검하고, 부족한 부분을 보완해야 한다. ISMS-P 인증을 보유한 기업은 AI 관련 통제 항목이 추가될 것에 대비해야 한다. N²SF 적용 대상 공공기관은 망분리 해제와 데이터 중심 보안으로의 전환 로드맵을 수립해야 한다. 고위험 AI 시스템을 운영 중인 기업은 안전성 평가와 모니터링 체계 구축을 준비해야 한다.

보안담당자의 역량 개발도 중요하다. AI 기술에 대한 기본적 이해 없이는 AI 보안을 담당할 수 없다. 머신러닝 기초, 대규모 언어모델의 작동 원리, AI 공급망 구조, AI 특화 공격 기법 등에 대한 교육이 필요하다. KISA가 운영하는 AI 보안 전문가 양성 과정이나 관련 자격증 취득을 고려할 만하다. 또한 법률·규제 측면에서도 EU AI Act, 미국 AI 행정명령, 국내 AI 관련 법안 동향을 지속적으로 모니터링해야 한다.

2025년이 보안담당자들에게 남긴 가장 큰 메시지는 명확하다. 보안은 더 이상 방어가 아니라 운영의 방식이라는 것이다. 기술이 아무리 발전해도 거버넌스와 책임체계가 없다면 하나의 인증키, 하나의 펨토셀, 하나의 배터리, 하나의 코드 실수로 수백만 명이 피해를 본다. AI 시대의 보안은 예방이 아니라 관리, 운영, 복원력, 책임이며, CISO는 이제 조직의 미래를 지키는 마지막 방어선이 아니라 첫 번째 설계자로 자리 잡아야 한다.

2026년, 우리는 무엇을 준비해야 하는가

2025년이 경고의 해였다면, 2026년은 설계의 해가 되어야 한다. 2026년 한국에서는 여러 흐름들이 본격적으로 정착될 것이다. 2025년까지 각 부처가 개별적으로 발표했던 AI 안전성 가이드라인이 통합되어 국가 차원의 안전성 표준이 등장할 것이다. 이 기준은 고영향 AI의 정의와 등급, 데이터·모델·추론·출력 전 과정의 위험요소, AI 사고보고와 감사 가능성, AI 설명가능성 범위, 편향과 환각 관련 안전 조치를 포함하게 된다. 2026년 한국 사회는 AI를 믿을 수 있는 사회를 위해 국가가 안전성의 문법을 제공하는 구조로 이동할 것이다.

공공기관 AI 도입 시에는 위험등급 평가, 데이터 민감도 분석, 복원력 평가, 책임자 승인, 운영 모듈 검증 같은 절차가 법·제도 수준으로 규정될 가능성이 높다. 한국의 공공 AI는 2026년에 속도보다 안전을 우선시하는 프레임으로 재정렬될 것이다. AI보안은 별도의 보안 분야로 존재하지 않고 보안 전체를 흡수하게 될 것이다. 기존의 네트워크 보안, 시스템 보안, 데이터 보안, 엔드포인트 보안, 앱 보안 등 세분화된 영역이 AI 기반 공격과 AI 기반 방어가 표준이 되면서 AI 중심으로 재편된다는 것이다.

또한 보안은 더 이상 기술 담당 부서가 아니라 경영과 운영의 언어가 된다. 기업의 이사회와 공공기관의 최고위층에서 AI 리스크가 기술 의사결정이 아니라 경영 의사결정이 되는 것이다. 2026년에는 국회에서 AI 안전성 관련 입법 논의가 더욱 활발해질 것이고, 기업 이사회에서 AI 리스크 보고가 필수화될 것이며, 공공기관에서 CAIO와 CISO가 공동 승인하는 의사결정 구조가 일반화되며, 금융·제조·의료 등에서 AI 사고 대응팀이 신설되는 장면을 어렵지 않게 보게 될 것이다.

또한 NIRS 화재 이후 한국의 공공 인프라는 대대적 재설계를 거치게 될 것이다. 하나의 데이터센터에 공공서비스를 몰아넣는 방식은 종말을 고하고, 재해복구와 업무연속성 확보를 위해 국가 디지털 인프라의 일정 비율은 민간 클라우드에서 운영될 가능성이 높다. 장애 예측, 리소스 자동 재할당, 인프라 자가치유 같은 기능을 가진 AI 기반 운영 플랫폼이 등장하고, 공공 인프라의 운영자 역할이 기술자에서 설계자로 이동할 것이다.

2026년 산업계가 직면하게 될 가장 중요한 변화는 AI 공급망 규제가 본격화된다는 점이다. AI SBOM 의무화, 데이터셋 출처 검증, 모델 학습 이력 로그 보존, 외부 API 및 모델 연동 승인제, 모델 재사용 시 위험기반 평가 도입 등이 제도화될 가능성이 높다. AI 기반 서비스가 늘어날수록 공급망의 투명성은 국가 경쟁력과 직결된다.

2026년의 기업과 공공기관에서는 CISO, CAIO, CPO 세 직책의 협업 체계가 표준이 될 것이다. CISO는 AI 보안과 리스크를 총괄하고, CAIO는 AI 도입과 운영을 총괄하며, CPO는 개인정보와 데이터 적법성을 총괄한다. 2025년까지는 이 셋이 분절적으로 움직여 왔지만, 2026년에는 AI 사업·보안·데이터 거버넌스가 서로 얽혀 있으며 세 직책이 단일한 위기 관리 체계를 구성해야 한다는 인식이 확산될 것이다.

국제적으로는 AI 규범 경쟁이 가속화될 것이다. 2026년 전 세계의 흐름은 AI 기술 경쟁이 아니라 AI 규범 경쟁으로 향할 것이며, 한국이 참여하거나 주도해야 할 핵심 분야는 AI 안전성 국제 표준, AI 공급망 보안 협력, 개인정보 이동 규범, AI 윤리·책임성 국제 협약, AI 기반 공격 대응 협의체 등이다. 2025년 사건들은 한국이 AI 기술 강국이기 이전에 AI 규범 강국이 되어야 함을 보여줬다.

경고에서 설계로, 2026년이 답할 차례

2025년은 한국 사회에 큰 충격을 남겼다. SKT 유심 해킹, KT 펨토셀 사고, NIRS 화재, 쿠팡 개인정보 유출 사건들은 하나의 공통된 메시지를 던진다. 한국 사회는 과거 기준으로는 설명할 수 없는 새로운 위험의 시대에 들어섰다는 것이다.

2025년은 충격과 경고의 해였다. 2026년은 그 경고를 기반으로 AI 시대에 적합한 국가적 보안·거버넌스·인프라 전략을 설계해야 하는 해다. AI 기본사회는 AI 안보사회 없이는 지속될 수 없고, 강건한 AI보안체계 없이 글로벌 AI 3강 실현도 없기 때문이다. 신뢰, 안전, 복원력, 책임성이 AI 사회의 토대이자 전제조건인 것이다. 우리가 2025년에 겪은 사건들은 강건한 AI 보안생태계로의 전환이 단순한 선택이 아니라 불가피한 국가적 과제임을 확인해 준 셈이다.

2025년 사건들의 공통점은 모두 전통적 보안모델이 상정하지 못한 영역에서 발생했다는 것이다. 내부 인증키 방치, 네트워크 설계 취약성, 물리적 인프라 재해 모두 기존 보안 모델이 중심적으로 다루던 취약점이 아니었다. 이제 위험은 네트워크 방화벽의 안쪽이나 바깥쪽으로 구분되지 않는다. 위험은 시스템을 둘러싼 설계, 운영, 공급망, 거버넌스, 사람, 절차 전체에 스며들어 있으며, 이러한 구조적 취약성이 AI 공격 시대에는 공격자의 전략적 자산이 된다. 따라서 AI 시대의 보안은 방어를 강화하는 것이 아니라 구조를 다시 설계하는 문제가 된다.

우리는 이제 AI에 대응하는 국가가 아니라 AI 시대를 설계하는 국가가 되어야 한다. 2025년의 사건들을 회고하며 우리는 이렇게 말할 수 있다. 한국은 AI 사회를 너무 빨리 시작한 것이 아니라, 우리는 AI 사회를 지킬 안전망을 충분히 갖추지 못한 채 시작했다고. 그러나 2026년 이후의 한국은 달라야 한다. AI 기술의 속도를 따라잡는 경쟁이 아니라, AI 시대의 위험을 예측하고, 관리하고, 통제하고, 그 위에서 AI를 신뢰할 수 있는 기술로 만드는 일을 해야 한다. 2025년에 우리는 AI 시대를 선언했지만, 이제는 AI 시대를 지킬 준비를 해야 한다는 것이다.

우리가 2025년에 겪은 충격은 동시에 기회이기도 하다. 한국은 빠르게 배우는 나라이고, 빠르게 대응하는 나라이고, 빠르게 제도를 재편하는 나라다. 2026년은 그 장점들이 가장 크게 발휘될 해가 될 것이다. 우리는 AI 시대를 선언했지만, 이제는 AI 시대를 지킬 준비를 해야 한다. AI에 적응하는 것이 아니라, AI 시대의 질서를 우리가 설계하는 국가가 되어야 한다. 기술의 속도보다 중요한 것은 그 기술이 서 있는 기반의 견고함이다. 그리고 2025년의 기록은, 그 기반을 다시 설계해야 한다는 사실을 분명히 보여줬다.

Part. 2

02 대규모 침해사고 발생과 침해사고 대응체계 개선

정보통신망법을 중심으로

강은성 교수 – 서울여자대학교 지능정보보호학부

올해는 유독 대규모 침해사고가 많이 발생하며 사회적으로 큰 문제가 됐다. 그에 따라 정부, 국회, 언론 등에서도 현황과 문제점을 파악하고, 대응 방안, 법적 · 제도적 개선책 마련에 대한 논의가 활발하다.

정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 정보통신망법)에서는 침해사고에 대한 정의와 침해사고 발생 시 정보통신서비스 제공자[1]의 신고 의무, 그에 대한 정부의 대응 책임 등을 관련 조항을 통해 침해사고에 대한 대응체계를 규정해 놓았는데, 올해 발생한 대규모 침해사고에서 이 규정들이 적합한지 의문이 제기되고 있다.
본 칼럼에서는 현 '침해사고 대응체계'의 현황과 한계를 살펴보고, 그에 대한 대안을 제시해 보고자 한다.

최근 침해사고 현황과 시사점

과학기술정보통신부(이하 과기정통부)에 따르면, 2021년 이후 침해사고 신고 건수는 매년 꾸준히 증가하는 추세다.

표 2-1 침해사고 신고 현황(2021년~2025년 상반기)

[단위 : 건수]

연 도	2021년		2022년		2023년		2024년		2025년
	상반기	하반기	상반기	하반기	상반기	하반기	상반기	하반기	상반기
건수	298	342	473	669	664	613	899	988	1,034
합계	640		1,142		1,277		1,887		1,034
증가율	N/A		78%		12%		48%		15%

– 2025년 증가율은 2024년 상반기 대비 2025년 상반기 증가율

출처: 과학기술정보통신부, 한국인터넷진흥원, 〈사이버 위협 동향 보고서〉, 2023~2025

1) 전기통신사업자와 영리를 목적으로 전기통신사업자의 전기통신역무를 이용하여 정보를 제공하거나 정보의 제공을 매개하는 자(정보통신망법 제2조(정의))

보고서에서는 사이버 위협 유형을 DDoS 공격, 악성코드(랜섬웨어), 서버 해킹, 기타(정보유출, 스팸 문자 및 메일 발송 등)로 나누는데, 주로 홈페이지 등 보안 관리가 취약한 중소기업을 대상으로 한 서버 해킹이 50% 이상을 차지한다.

하지만, 일반 국민이 체감하는 침해사고의 문제는 숫자로 보는 것 이상이다. 해킹을 통한 SKT의 약 2,700만 명의 유심(USIM) 정보 등 고객정보 유출, 랜섬웨어 공격으로 인한 국내 최대 온라인서점 예스24의 서비스 마비(1차 6월, 2차 8월)와 시민들의 필수 서비스인 SGI서울보증의 보증 서비스 중단(7월), 해킹에 의한 롯데카드 297만 명 고객정보 유출, 이동통신장비 관리 부실로 인한 kt 무단 결제(9월), 프랙(Phrack) 보고서로 알려진 LG유플러스 계정정보 유출(10월), 인증정보 관리 부실로 발생한 퇴직자에 의한 쿠팡의 3,370만 명 고객정보 유출(11월) 등 일반 국민과 밀접하게 관련된 서비스에서 대규모 침해사고가 발생하면서 사회적인 이슈가 됐다.
정보통신망법의 관련 규정에 따르면, 침해사고가 발생하면 그에 대한 조사 · 분석은 다음과 같이 4단계로 이뤄진다.

표 2-2 **침해사고 조사 · 분석 단계**

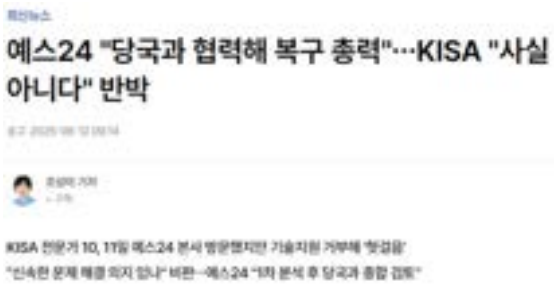
단계	활동	주체
1	침해사고 신고	사고기업
2	침해사고 조사 · 분석에 대한 동의	사고기업
3	침해사고 조사 · 분석 및 재발방지책 수립	한국인터넷진흥원(KISA)
4	이행 명령	과기정통부

LG유플러스 사건은 침해사고 신고 단계(1단계)의 문제를 드러냈다. 올해 8월 프랙 보고서에서 이 회사의 협력업체를 통해 계정정보가 유출된 것이 증거와 함께 공개됐으나, 회사는 이미 7월에 관련 제보를 받고 자체 조사를 한 뒤 침해사고가 발생한 정황이 없다면서 버티다가 국회 국정감사에서 압박을 받은 CEO가 신고하겠다고 한 뒤에 비로서 침해사고를 신고한 것이다.

이후 민관합동조사단을 꾸려 조사한 과기정통부는 12월, "개인정보 유출 관련 서버의 고의 폐기 의혹"이 있다며 수사를 의뢰했는데, 이게 사실로 밝혀지면, 침해사고 증거를 은닉 또는 인멸하기 위해 신고를 지연시킬 수도 있다는 우려가 현실로 드러난 사건이 될 것으로 보인다.[2)]

예스24 사건은 침해사고 신고가 접수됐다 하더라도, 사고기업의 동의(2단계) 없이는 정부가 조사 · 분석을 하지 못하는 현행 법 체계의 한계를 보여줬다.

2) MBC, "[단독] 과기부 'LGU+ 개인정보 유출 관련 서버 고의 폐기 의혹 수사의뢰'", 2025.12.10, https://imnews.imbc.com/news/2025/econo/article/6783846_36737.html


예스24 "당국과 협력해 복구 총력"…KISA "사실 아니다" 반박

(출처: 연합뉴스 2025.6.12)

예스24는 "한국인터넷진흥원과 협력해 복구에 총력을 다하고 있다"고 밝혔으나, KISA는 분석가들이 두 차례 방문했으나, 거부당했다고 이례적으로 보도자료를 낸 것이다.[3] 예스24는 여론의 뭇매를 맞고서야 협조함으로써 사고기업에 대한 조사 · 분석의 강제성이 필요하다는 의견이 공론화되는 계기가 됐다.

침해사고 대응체계의 현황과 문제점

정보통신망법에 침해사고 대응에 관한 규정이 들어온 것은 슬래머웜으로 전국의 인터넷이 마비된 '1.25 대란'이 발생한 다음 해인 2004년의 일이다. 이후 20여 년이 흘러 지나 IT 인프라에 문제가 생기면, 국민이 일상생활에서 큰 어려움을 겪을 정도로 'IT 기반 사회'가 되어 침해사고 대응의 중요성은 더욱 커졌다.

현행 정보통신망법에서 규정한 침해사고 대응에 관한 과기정통부장관의 업무는 다음 4가지로 요약된다.

(1) 침해사고에 관한 정보의 수집 · 전파, 침해사고의 예보 · 경보, 침해사고에 대한 긴급조치, 그 밖에 대통령령으로 정하는 침해사고 대응조치(법 제48조의2(침해사고의 대응 등) 제1항, 시행령 제56조(침해사고 대응조치))

(2) 침해사고 원인 분석, 피해 확산 방지, 사고대응, 복구 및 재발 방지를 위한 대책을 마련, 사고기업에 이행 명령, 이행 명령에 대한 이행 여부 점검 및 보완 필요 시 시정 명령(법 제48조의4(침해사고의 원인 분석 등) 제2항, 제3항)

(3) 침해사고 분석에 필요한 정보통신망 접속 기록 등 관련 자료의 보전 및 제출 명령(법 제48조의4(침해사고의 원인 분석 등) 제5항, 제6항)

(4) 중대한 침해사고 발생 시 원인 분석을 위해 민관합동조사단의 구성, 운용(법 제48조의4(침해사고의 원인 분석 등) 제4항)

3) 한국인터넷진흥원, "KISA, 예스24 2차 입장문에 대한 설명", 2025.6.11, https://www.kisa.or.kr/402/form?postSeq=2508

현행 침해사고 대응 체계의 문제점을 정리하면 다음과 같다.

첫째, 모든 침해사고에 대해 신고 요건과 제재가 동일하다는 점이다.

정보통신망법에서는 침해사고와 침해사고 발생에 따른 신고 의무를 다음과 같이 규정한다.

법 제2조(정의)

7. "침해사고"란 다음 각 목의 방법으로 정보통신망 또는 이와 관련된 정보시스템을 공격하는 행위로 인하여 발생한 사태를 말한다.

가. 해킹, 컴퓨터바이러스, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법

나. 정보통신망의 정상적인 보호 · 인증 절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등을 정보통신망 또는 이와 관련된 정보시스템에 설치하는 방법

법 제48조의3(침해사고의 신고 등) ① 정보통신서비스 제공자는 침해사고가 발생하면 즉시 그 사실을 과학기술정보통신부장관이나 한국인터넷진흥원에 신고하여야 한다. 이 경우 정보통신서비스 제공자가 이미 다른 법률에 따른 침해사고 통지 또는 신고를 했으면 전단에 따른 신고를 한 것으로 본다.

시행령 제58조의2(침해사고 신고의 시기, 방법 및 절차) ① 정보통신서비스 제공자는 법 제48조의3 제1항 전단에 따라 침해사고를 신고하려는 경우에는 침해사고의 발생을 알게 된 때부터 24시간 이내에 다음 각 호의 사항을 과학기술정보통신부장관 또는 한국인터넷진흥원에 신고해야 한다.

침해사고의 발생 일시, 원인 및 피해내용

침해사고에 대한 조치사항 등 대응 현황

침해사고 대응업무를 담당하는 부서 및 연락처

법 제76조(과태료) ① 다음 각 호의 어느 하나에 해당하는 자와 제7호부터 제11호까지의 경우에 해당하는 행위를 하도록 한 자에게는 3천만 원 이하의 과태료를 부과한다.

6의6. 제48조의3 제1항을 위반하여 침해사고의 신고를 하지 아니한 자

정보통신망법의 침해사고 정의에 따르면, PC에서 악성코드를 발견하더라도 침해사고 신고를 해야 하고, 이를 위반하면 3000만 원 이하의 과태료를 부과받을 수 있다. 극단적으로 비교하면, 중소기업 PC가 악성코드에 감염된 침해사고와 매출액 40조 원의 기업에서 3300만 명의 고객정보가 유출된 침해사고에 대해 신고 의무와 제재 기준이 같다. 이런 정도의 제재라면 후자의 경우 해당 대기업에 미신고 또는 지연 신고의 동기가 생길 수 있지 않을까 싶다.

또한, 법에서는 24시간 이내에 ▲침해사고 발생 일시, ▲원인 및 피해 내용, ▲침해사고 대응 현황 등을 예외 없이 작성하도록 하고 있다. 숙련된 정보보안 인력을 갖추고 있지 못한 기업에는 좀 벅차 보이는 요건으로 보인다.

둘째, 사고기업의 신고와 동의 없이는 침해사고 대응 활동이 이뤄지기 어렵다.

세계 최고 수준의 IT 인프라를 갖춘 우리나라에서는 IT 인프라가 마비되면 많은 국민의 일상이 매우 불편해지거나 심지어 일부는 멈춰 서기도 한다. 공격자는 취약한 사이트를 공격하고 이를 경유지로 다른 사이트를 공격한다. 침해사고로 개인정보가 대량으로 유출되면 국민의 삶이 위협받는다.

이러한 상황에서 과기정통부장관은 신속한 조사 · 분석을 통해 침해사실을 파악하고, 원인을 분석해 피해 확산을 차단하는 등 법에서 정한 침해사고 대응 업무를 수행해야 하는데, 현 대응체계에서는 사고기업의 신고와 조사에 대한 동의가 없으면 이를 수행하기 어렵다.

이것은 침해사고 신고율이 낮다는 점에서 더 큰 문제로 지적된다.

그림 2 – 1 정보 침해사고 신고 및 미신고 이유(1+2순위) 대상 기간 2023년 1월~12월

출처: 한국정보보호산업협회, 〈2024 정보보호 실태조사〉, 2024.12.

〈2024 정보보호 실태조사〉(한국정보보호산업협회, 2024)에 따르면, 2023년 한 해 동안 침해사고를 경험한 기업 중 신고한 기업은 19.6%에 불과하다. 이것도 직전 연도에 비해 11.1%나 늘어난 것이다. 미신고 이유 중 가장 많은 답변은 '경미한 침해사고'(73.7%)인데, 일반적으로 '경미한 침해사고'라고 하면 짧은 시간의 DDoS 공격, 단순 악성코드 감염 등 피해가 작고, 분석의 필요성이나 중요도가 떨어진 사고를 말한다. 침해사고 전체를 파악하는 데 의미를 부여할 수 있으나, 분석 대상으로서 의미는 적을 수 있다.

그런데 같은 보고서에서 경험한 침해사고 유형에 대한 답변(복수 응답)은 이와는 좀 다르다.

〈그림 2〉에 따르면, 경험한 침해사고 유형의 1~3위가 ▲랜섬웨어 감염, ▲해킹, ▲사이버 공격으로 인한 IT시스템 마비인데, 이 유형들은 '경미한 침해사고'에 포함되기는 어려워 보인다. 오히려 〈그림 1〉의 미신고 이유 중 '신고 업무 복잡'(54.3%), '피해 사실 공개 우려'(17.6%)에 포함될 가능성이 높다. '신고 업무 복잡'에는 24시간 이내에 침해사고 발생 일시, 원인 및 피해 등을 작성하기 어려운 것이 포함됐을 가능성이 있다. 앞에서 예를 든 kt나 LG유플러스의 미신고(또는 지연 신고) 이유는 '피해 사실 공개 우려'에 해당할 것으로 보인다.

그림 2 – 1 경험한 침해사고 유형

출처: 한국정보보호산업협회, 〈2024 정보보호 실태조사〉, 2024.12.

셋째, KISA의 침해사고 분석 인력이 부족하다.

사실 KISA의 침해사고 분석 인력의 부족 문제는 어제오늘의 일이 아니다. 이미 2017년에 KISA가 침해사고에 대해 '사이버 보건소' 역할을 해야 하므로, 조직과 인력을 대폭 확대해야 한다는 주장이 제기된 적이 있으나, 별로 호응을 얻지는 못했다.[4)]

지금도 KISA 홈페이지 조직도에서 찾아보면 침해사고 분석 인력은 채 20명이 되지 않는다. 민관합동조사단이 꾸려질 정도의 큰 사건이 발생하면, KISA 인력은 몇 주, 몇 달 동안 야근과 주말 근무를 하는 때도 있는데, 담당 인력의 업무 부하는 많이 늘어날 수밖에 없다.

이 문제에 대처하기 위해 KISA는 침해사고 분석 대상 사건의 일부를 다시 침해사고 분석 전문기업에 위탁해 처리하지만, 여전히 침해사고 분석 업무에 비해 인력이 부족하다는 것이 일반적인 진단이다. 전문기업 역시 KISA의 예산이 한정되어 있고, KISA가 발주하는 용역만을 바라보고 사업을 할 수는 없기 때문에 인원을 마냥 늘리기는 쉽지 않은 게 현실이다.

넷째, CTI의 축적과 공유에 관한 문제다.

정보통신망법에 침해사고 대응 관련 규정이 신설된 2004년부터 침해사고 신고처이자 조사 · 분석 전문기관이 된 KISA는 20년 이상 쌓아온 기술력과 방대한 분석 데이터를 보유하고 있다. 이를 관리하고, 분석 · 가공해 필요한 정보를 민간에 공유함으로써 우리 사회의 사이버 위협에 대응하는 것이 KISA가 가진 또 하나의 역할이다.

4) 아이뉴스24, "'사이버 보건소' 되겠다는 KISA", 2017.7.9, https://n.news.naver.com/mnews/article/031/0000417034?sid=105

KISA는 KISA 보호나라 사이트에서 '보안 공지'를 통해 보안취약점이나 보안 패치 관련 사항을 공지하고[5], '사이버 위협정보 분석 · 공유시스템' C-TAS(Cyber Threat Analysis & Sharing)[6]를 통해 주요 침해사고에 대한 정보를 신속하게 공유하여 기업들의 침해사고 사전 · 사후 대응에 많은 도움을 주고 있다.

KISA가 제공한 본격적인 CTI 자료로는 주요 사이버 위협을 분석한 'TTPs #N' 사이버 위협 분석 보고서 시리즈가 있다.[7] 이 보고서에서는 주요 침해사고에서 발견된 사이버 위협을 MITRE ATT&CK 기반으로 심도 있게 분석했는데, 보고서마다 조회 수가 수만 회에 이를 정도로 '보안동네'에서는 인기 있는 문서였다.

그림 2 – 1 사이버 위협 분석 보고서 목록

출처: KISA 보호나라

다섯째, 최고급 인력이 투입된 침해사고 분석 서비스가 '무료'로 제공된다.

적지 않은 최고의 침해사고 분석 전문가들이 투입되어 사고기업에 발견하지 못한(않은?) 사고의 원인을 찾아내고 분석해 주는 서비스가 굴지의 대기업에도 무료로 제공된다. 침해사고가 일어난 국내 대기업들이 종종 이용하는 맨디언트(Mandiant)에게 비슷한 서비스를 받았다면 아마도 수억 원은 냈을 것이다.

(세금이 투입되는 거라서 실제로 무료는 아닌데, 자금 여력이 있는 중견기업 이상에서 발생한 침해사고 분석을 위해 세금을 사용하는 것이 적절한지는 의문이다.)

일정 규모 이상의 기업에는 유료로 제공하고, 고생한 분석가들에게 인센티브로 지급하는 것도 고려할 만하다. 일부는 침해사고 대응 교육이나 중소기업 침해사고 대응에 활용할 수 있을 것 같다.

5) KISA보호나라&KrCERT/CC 보안 공지, https://boho.or.kr/kr/bbs/list.do?menuNo=205020&bbsId=B0000133
6) C-TAS, https://ctas.krcert.or.kr/
7) KISA TTPs 침해사고 분석 보고서, https://www.boho.or.kr/kr/bbs/list.do?searchCnd=1&bbsId=B0000127&searchWrd=TTP&menuNo=205021

침해사고 대응체계 개선을 위한 제언

이러한 문제의식을 바탕으로 침해사고 대응체계 개선을 위해 다음 몇 가지를 제안한다.

첫째, 침해사고 신고 의무 위반에 대해 제재는 차등 적용되는 것이 바람직하다.

법에서 부과한 침해사고 신고 의무 위반 시 기업과 서비스의 규모나 중요성, 침해사고의 중대성에 따라 제재가 달라지는 것이 합리적이다. 수조 원의 매출을 올리는 서비스를 운영하는 대기업이 해킹을 당해 수천 만 명의 고객정보가 유출된 상황에서 신고 의무를 위반했는데, 3000만 원 이하의 과태료(그것도 1회 위반 시 750만 원)를 부과하는 것은 제재로서 의미가 없다. 이러한 행위에 대해서는 수백억 원의 과징금을 부과할 수 있어야 한다. 특히, 증거를 은닉 · 인멸하려는 목적이 있었다면, 일반적인 신고 의무 위반 때보다 더 많은 과징금을 부과하는 것이 타당하다.

이에 관해서는 유럽연합의 '네트워크 및 정보시스템 보안2'(NIS2: Network and Information System 2) 지침을 참고할 만하다.[8] NIS2 지침은 공공기관과 기업의 규모, 산업 섹터를 종합하여 필수 기관과 중요 기관을 정의하는데(제3조(필수 기관과 중요 기관)), 이들 기관에 중대 사고가 발생하면 서비스 이용자에게 지체 없이 통지하도록 하고, 사고를 인지한 지 24시간 이내 조기 경보, 72시간 이내 사고 통지(심각도와 영향을 포함한 초기 평가, 가능하면 침해지표(IoC) 포함), 1달 이내에 최종 보고서를 국가 CSIRT(Computer Security Incident Response Team) 또는 규제 기관에 제출하도록 했다(제23조(보고 의무)).

보고 의무를 위반하면, 필수 기관에는 최소 1000만 유로(약 170억 원) 또는 직전 회계연도 연간 전세계 총 매출액의 최소 2% 중 더 높은 금액의 과징금을, 중요 기관에는 최소 700만 유로(약 120억 원) 또는 직전 회계연도 연간 전세계 총 매출액의 최소 1.4% 중 더 높은 금액의 과징금을 부과할 수 있다.

예를 들어, 필수 기관이 중대 사고를 중대 사고가 아니라고 판단해 보고하지 않거나 지연 보고하면 이 조항을 위반한 것으로 간주되어 고액의 과징금을 부과받을 수 있다.

추가적으로 신고 내용에 대한 예외 규정이 필요해 보인다. 시행령 제58조의2(침해사고 신고의 시기, 방법 및 절차) 제1항에 따르면, 사고기업은 침해사고 발생 시 24시간 이내에 ▲침해사고의 발생 일시, 원인 및 피해 내용, ▲침해사고 대응 현황, ▲담당 부서 및 연락처를 작성해 신고해야 한다. 하지만, 24시간 안에 이걸 모두 파악하여 작성하는 일은 쉽지 않다. 예를 들어, 24시간 이내에는 발생 사실과 담당 부서 및 연락처, 그리고 신고해야 할 사항 중 그때까지 파악한 사항을 신고하고, 다른 사항은 파악한 시점에 신고하게 하는 등의 예외 규정을 두는 것이 좋겠다.

8) 유럽연합 NIS2 지침 https://eur-lex.europa.eu/eli/dir/2022/2555/oj. 발효되면 즉시 유럽연합 전체에 적용되는 규정(Regulation)과는 달리 지침(Directive)은 회원국에서 국내법에 반영해야 해당 국가에서 법적 효력이 발생한다.

개인정보 유출 시 통지 및 신고 의무를 부과한 개인정보보호법에서는 이와 비슷한 규정이 있다.

개인정보보호법 시행령 제40조(개인정보 유출 등의 신고)
(…중략…)
② 제1항에도 불구하고 개인정보처리자는 제1항에 따른 신고를 하려는 경우로서 법 제34조(개인정보 유출 등의 통지 · 신고) 제1항 제1호 또는 제2호의 사항에 관한 구체적인 내용을 확인하지 못한 경우에는 개인정보가 유출 등이 된 사실, 그때까지 확인된 내용 및 같은 항 제3호부터 제5호까지의 사항을 서면 등의 방법으로 우선 신고해야 하며, 추가로 확인되는 내용에 대해서는 확인되는 즉시 신고해야 한다.

즉, 사고기업(개인정보처리자)은 개인정보가 분실 · 도난 · 유출(이하 유출 등)이 됐음을 알게 되었을 때 72시간 이내에 (1)유출 등이 된 개인정보 항목, (2)시점과 경위, (3)피해 최소화를 위해 정보주체가 취할 방법, (4)사고기업의 대응 조치, (5)담당 부서 및 연락처를 규제 기관에 신고해야 하는데, 이때 (1), (2)를 확인하지 못하면, (3)~(5)를 먼저 신고하고, (1), (2)는 확인될 때 신고하는 예외 규정을 뒀다.

둘째, 적절한 요건을 갖추면 사고기업의 신고나 동의 없이도 침해사고를 조사 · 분석할 수 있도록 할 필요가 있다.
예를 들어, 침해사고의 정황 또는 침해사고와 연관된 피해 발생 정황이 드러난 경우나 긴급 대응이 필요하다고 판단되면, 사고기업의 신고나 동의 없이도 조사 · 분석을 할 수 있게 하는 것이다.

지난 10월, 과기정통부와 관계부처가 발표한 '범부처 정보보호 종합대책'에서는 "해킹 정황을 확보한 경우에는 기업의 신고 없이도 정부가 신속히 현장을 조사할 수 있도록 정부의 조사 권한을 확대"한다고 발표했다.[9] 이러한 내용을 포함해 여야가 합의한 정보통신망법 개정안이 국회 본회의 통과를 앞두고 있다.[10]

셋째, 사고기업의 원인 분석 의무를 강화할 필요가 있다.
정보통신망법에서는 침해사고가 발생한 기업의 의무로 ▲침해사고 신고 ▲침해사고 원인 분석 ▲피해 확산 방지 ▲복구 및 재발 방지 ▲과기정통부장관의 요구에 따른 자료 보전 및 제출 등을 적시한다(법 제48조의3, 제48조의4).
그런데 그동안 신고 의무는 강조됐으나, 침해사고 원인 분석과 그에 따른 조치 의무는 그리 강조되어 오지 않은 것 같다. 의무 규정인데, 위반에 따른 제재도 없다.

9) 과기정통부, "범부처 정보보호 종합대책 발표", 대한민국 정책브리핑, 2025.10.22. https://www.korea.kr/briefing/pressReleaseView.do?newsId=156722945

10) [2214896] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(대안)(과학기술정보방송통신위원장), 국회에 제출된 관련 개정안들이 통합된 과학기술정보방송통신위원장 대안이 법제사법위원회를 통과하여 특별한 일이 없으면 국회 본회의를 통과할 것으로 보인다(2025.12.17. 기준).

법 제48조의4(침해사고의 원인 분석 등) 제1항은 다음과 같다.

> **법 제48조의4(침해사고의 원인 분석 등)** ① 정보통신서비스 제공자 등 정보통신망을 운영하는 자는 침해사고가 발생하면 침해사고의 원인을 분석하고 그 결과에 따라 피해의 확산 방지를 위하여 사고대응, 복구 및 재발 방지에 필요한 조치를 하여야 한다.

기업에서는 침해사고가 발생하면 어떤 식으로든 조사 · 분석을 진행해 경영진에 보고하는 것이 보통이다. 침해사고 대응 전문업체나 보안관제업체에 위탁하거나 스스로 전문역량을 갖추고 있다고 판단하면 자체적으로 수행하기도 한다.

법적으로 사고기업에서 수행한 사고 원인 분석을 보고서로 작성해 과기정통부장관(또는 KISA)에 제출하게 할 필요가 있다. KISA에서는 제출 받은 보고서 내용을 검증해 보완을 요구할 수 있어야 한다. 법 제48조의4(침해사고의 원인 분석 등)에서 과기정통부장관에 침해사고 원인 분석부터 재발 방지책 수립, 사고기업에 이행 명령의 책임과 권한을 부여한 취지에도 부합한다.

이를 통해 사고기업은 침해사고 원인을 깊이 있게 이해해 정보보안 역량을 강화하고, 재발 방지책을 구현하는 등 기업의 전반적인 보안 수준을 높이는 계기가 될 수 있다. 또한, 국가적으로는 CTI를 축적 · 관리 · 공유함으로써, 침해사고 예방을 위해 사용되는 선순환이 이뤄질 것으로 기대된다.

보고서 제출 의무는, 1차적으로 ISMS 인증의무대상자, 개인정보 등 중요 고객정보가 유출된 정보통신서비스 제공자, 본인확인기관, 주요정보통신기반시설 관리기관(정보통신기반보호법), 전자서명인증사업자(전자서명법), 금융회사 및 전자금융업자(전자금융거래법) 등 정보통신망법이나 다른 법에서 지정된 기관 중 정보보안이 중요한 기관에만 부과하고, 필요 시 이후 확대해 나가는 것이 바람직하다. 여기에 포함되지 않는 사고기업이 보고서를 제출하면 그에 대한 인센티브를 부여하는 방법도 고려할 수 있다.

보고서 제출 의무 위반에 따른 제재 규정도 추가할 필요가 있다. 다만, 보고서를 통해 기업의 내부 정보가 유출되면 사고에 대한 손해배상 소송 등에 불리할 수 있다는 우려가 있을 수 있으므로, 그에 대한 대책을 보완하는 것이 필요하다.

넷째, KISA 침해사고 분석 인력을 확충해야 한다.

올해 대규모 침해사고가 여럿 발생함에 따라 그 어느 때보다 KISA의 침해사고 분석 인력을 확충해야 한다는 공감대가 형성돼 있다. 특히, 침해사고 신고 의무 위반에 대한 제재가 강화되고, 사고기업의 신고나 동의 없이 침해사고 분석이 가능해지면, 분석 대상이 되는 침해사고가 더 증가할 것으로 예상된다. KISA의 침해사고 분석 인력이 확충돼야 할 이유다. 다만, KISA에서 경험을 쌓은 주요 분석 인력이 민간기업으로 이직하는 사례도 종종 있어서 그에 대한 검토와 대책 수립 또한 필요한 것으로 판단된다.

다섯째, KISA가 CTI를 축적 · 관리 · 공유하는 체계를 갖추는 것이 좋겠다.

20년 이상의 사고 분석을 통해 확보한 자료를 안전하게 보관하고, 이용하기 쉽게 관리하며 적절한 방식으로 공유하면 우리 사회의 보안 수준을 높이는 데 상당히 도움이 될 것으로 판단된다.

'보안 동네'에는 국내외 보안업체에서 자체 제품을 판매하기 위해 제공하는 관련 기술에 관한 자료는 여럿 있으나, 체계적이고 객관적이며 깊이 있는 자료를 찾아보기는 쉽지 않다. KISA에서 생산한 'TTPs #N' 보고서가 인기 있던 이유다. 인력과 시스템을 확보하여 KISA가 계속 깊이 있는 CTI를 공급해 주기를 바란다. 필요하면, 정보통신망법에 이에 관한 KISA의 역할을 적시하는 방법도 검토해 보면 좋겠다.

여섯째, 전문적인 침해사고 분석이 가능한 기업이 많이 생길 수 있도록 제도를 갖춰야 한다.

침해사고 분석을 위해서는 역량과 경험을 갖춘 전문인력뿐 아니라 분석 체계와 분석 도구, 관련 조직이 필요한데, 이 시장이 제대로 형성되어 있지 않은 국내에서 이러한 요건을 갖추기가 쉽지 않다. 당연히 분석 역량을 갖춘 사고기업이 그리 많지 않고, 어느 보안기업이 분석 역량을 갖췄는지 알기도 어렵다. 보안관제업체는 대부분 분석팀을 보유했으나, 침해사고 분석 역량 기준으로 본다면 편차가 있는 것으로 보인다. 전문역량이 부족한 조직이 침해사고를 분석하면서 데이터를 오염시키는 사고가 발생할 수 있다는 주장도 있다.

이를 위해 '침해사고 분석 전문기업 지정제도'가 필요하다. 전문기업이 지정되면, 사고기업은 신뢰할 수 있는 전문기업에 의뢰해 사고 원인을 분석하고 재발 방지책을 수립, 적용할 수 있으며, 과기정통부 역시 민관합동조사단 구성이나 대규모 침해사고 대응 시 지정 전문기업과 협업할 수 있다. KISA가 지정 기업과는 CTI 공유 등 좀 더 긴밀한 협업을 진행할 수도 있을 것이다.

이와 비슷한 제도로, 영국에서는 NCSC(National Cyber Security Centre)에 'Assured Cyber Incident Response(CIR) provider'가 50여 개 등록되어 있고[11], 유럽연합 역시 회원국 CSIRT를 중심으로 적절한 절차를 거쳐 침해사고 대응 전문기업 풀을 유지한다. 2025년 2월에 발효된 유럽연합 사이버연대법(Cyber Solidarity Act)에서는 대규모 침해사고 발생 시 유럽연합 차원에서 대응할 수 있도록 '사이버 보안 예비군'(Cybersecurity Reserve) 설립을 적시했다.[12]

이미 과기정통부에서는 정보보호 전문서비스기업 지정제도와 보안관제 전문기업 지정제도를 시행하고 있는데, 전자는 정보통신기반보호법에 따른 "주요정보통신기반시설의 취약점 분석 · 평가 업무 및 보호대책 수립업무"라는 시장이 있고, 후자는 국가사이버안전관리규정에 따른 "국가 공공기관 보안관제 센터 운영"이라는 시장이 있는 데 비해 '침해사고 분석 시장'이 안정적으로 만들어질 수 있는지 우려가 제기되기도 한다.
시장의 수요를 형성한다는 관점에서는 앞서 제안한 사고기업의 침해사고 분석 및 제출의무가 출발점이 될 수 있을 것으로 보인다. 공급 관점에서는 KISA의 무료 '침해사고 분석 서비스'를 서서히 줄여 나갈 필요가 있다. KISA의 전문인력은 CTI 공급과 위협 헌팅, 공격자 추적 등 민간기업에서 하기 어려운 국가 차원의 일에 집중하는 것이 어떨까 생각된다.

11) NCSC, "Cyber Incident Response", https://www.ncsc.gov.uk/schemes/cyber-incident-response/find-a-provider
12) 유럽연합 사이버연대법 https://eur-lex.europa.eu/eli/reg/2025/38/oj

다만, 앞서 언급한 두 지정제도의 전문기업 매출도 법규에서 정한 영역에서만 나오는 것은 아니다. 전문기업 지정을 통해 얻은 전문성과 신뢰성을 바탕으로 다른 영역으로 시장을 확대하는 것이 보통이다. 유럽의 CIR 지정업체의 경우도 마찬가지다.

글을 마치며

올해 대규모 침해사고가 여러 건 발생하면서 일반 국민들의 불안이 커졌다. 침해사고를 100% 막기 어려우니 어쩔 수 없다는(주로 보안 분야 종사자의) 주장이 있는 한편, 그 정도 매출을 올리면서 침해사고를 예방하지 못하느냐는(일반 국민의) 비난도 상당하다. 모든 사업자가 침해사고를 당하는 것은 아니고, 보유한 정보자산과 예방–탐지–대응의 수준에 따라 피해 차이도 상당해서 두 주장의 사이 어디쯤 우리가 나아갈 좌표가 있을 것 같다.

수백 년을 중앙집권국가로 살아온 우리나라에서는 큰 사고가 일어났을 때 정부에 대한 비판 여론이 비등하면 다소 급하게 법령 · 정책 · 제도가 만들어지곤 한다. 이따금 이렇게 만들어진 법령 · 정책 · 제도가 현장에서 제대로 작동하지 않거나 오히려 역으로 작동하기도 한다. 2008년, 2011년, 2014년에 이어 2025년이 그런 해가 될지도 모르겠다.

정보통신망법에서는 침해사고의 예방과 대응을 위한 정보통신서비스 제공자와 과기정통부의 역할과 책임을 규정한다. 이 글에서는 정보통신망법을 중심으로 침해사고 대응체계의 현황과 문제점을 짚어 보고, 대안을 다뤘다. 해 아래 새 것이 없듯이 이미 여러 곳에서 제안된 것이나 준비가 진행되는 것도 있을 것이다. 다양한 의견이 민–관–학–연 전문가들의 충분한 논의와 일반 국민들의 공감을 거쳐 실효성 있는 법령 · 정책 · 제도가 만들어지기를 바란다.

Part. 2

03 「범부처 정보보호 종합대책」, 제도 개선 방향과 과제

손경민 변호사 – 법무법인(유) 광장

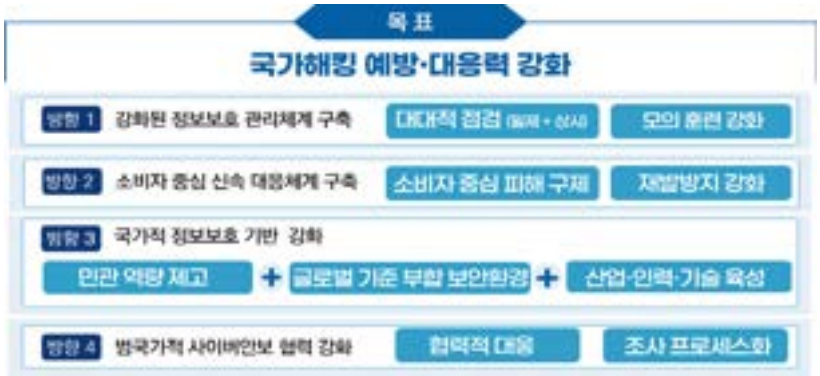
관계부처(과학기술정보통신부, 기획재정부, 금융위원회, 개인정보보호위원회, 행정안전부 등)는 2025. 10. 22. 「범부처 정보보호 종합대책」(이하 "정보보호 종합대책")을 발표했다.

2025년에는 통신, 금융, 전자상거래 등 분야를 막론하고 해킹사고가 반복되어 발생했는데, 정부는 이러한 일련의 해킹 사고를 심각한 위기 상황으로 인식하고, 범정부 차원의 유기적인 대응 체계를 즉시 가동하기 위한 것이다. 그리고 국가 전반의 정보보호 역량을 강화하려는 것을 목표로 제시했다.

정보보호 종합대책은 현 사안의 시급성을 고려하여 즉시 실행할 수 있는 단기과제 위주로 제시했으며, 이후 중장기 과제를 망라하는 「국가 사이버안보 전략」을 연내 수립할 계획이라는 점도 선제적으로 밝혔다.

정보보호 종합대책의 주요 추진 방향은 (1) 국민 생활에 밀접한 핵심 IT 시스템의 대대적인 보안 점검을 추진, (2) 소비자 중심의 사고 대응 체계 구축과 재발 방지 대책의 실효성을 강화, (3) 민 · 관 전반의 정보보호 역량을 강화하는 한편, 글로벌 기준에 부합하는 정보보호 환경 조성과 정보보호 산업 · 인력 · 기술을 육성, (4) 범국가적 사이버안보 협력 체계 강화를 그 내용으로 한다.

그림 3 – 1 범부처 정보보호 종합 대책의 주요 추진 방향



정보보호 종합대책의 주요 내용은 이후 관계부처의 보도자료 등 정책자료, 국회 법률안 등을 통해 구체화되고 있다. 즉, 정보보호 종합대책은 현재 정책적 선언을 넘어 구체적인 법률과 제도로 구현되는 단계에 접어들고 있다. 그에 따라 향후 정보보호 분야의 법률 및 제도에 중대한 변화를 가져올 것으로 예상된다. 이러한 시점에서 동 대책의 내용을 정확히 이해하는 것은 향후 제도 변화의 방향을 예측하고 이에 대비하기 위해 필요할 것이다. 현재 진행 중인 제도 개선 논의와 앞으로의 변화를 선제적으로 살펴보겠다.[1)]

1. 강화된 정보보호 관리체계 구축

가. 핵심 IT 시스템에 대한 점검

해킹에 대한 국민들의 만연한 불안감 해소를 위해, 공공 · 금융 · 통신 등 국민 대다수가 이용하는 1,600여개 IT 시스템들에 대해 대대적인 보안 취약점 점검을 즉시 추진하기로 했다. 대상 시스템에는 공공기관 기반시설 288개, 중앙 · 지방 행정기관 152개, 금융업 261개, 통신 · 플랫폼 등 ISMS 인증기업 949개 등이 포함된다.

특히 통신사의 경우에는 실제 해킹 방식의 강도 높은 불시 점검을 추진하고 주요 IT 자산에 대한 식별 · 관리체계를 구축하도록 한다. 아울러 소형기지국(펨토셀)은 안정성이 확보되지 않을 경우 즉시 폐기하는 등 보다 엄격히 조치할 계획임을 밝혔다.

그에 따라 과학기술정보통신부(이하 “과기정통부”)는 통신, 온라인쇼핑몰 등 900여개 ISMS 인증기업들을 대상으로 모든 인터넷 접점에 대한 보안 취약점 점검 등 긴급 자체 점검을 실시하도록 요청했으며, 기업들의 점검 결과에 대해 내년 초부터 현장 검증을 실시할 예정이다.

나. 보안 인증 실효성 강화

정보보호 관리체계(ISMS) 인증, 정보보호 및 개인정보보호 관리체계(ISMS-P) 인증의 실효성을 전면 강화하기로 했다. 현장 심사 중심으로 전환하고 중대한 결함이 발생할 경우 인증을 취소하는 등 실효성을 제고하고 사후관리를 강화하겠다는 것이다.

그에 따라 개인정보보호위원회(이하 “개인정보위”)와 과기정통부는 2025. 12. 6. “정보보호 및 개인정보보호 관리체계 인증 실효성 전면 강화” 보도자료를 발표하였다. 주요 개선 계획은 아래와 같다.

(1) ISMS-P 인증 의무화: 기존 자율적으로 운영되던 ISMS-P 인증을 공공 · 민간 주요 개인정보처리시스템(주요 공공시스템, 통신사, 대규모 플랫폼 등)에 대해 의무화

1) 본 기고문은 2025. 12. 15.에 작성되었음

(2) 강화된 인증기준 적용: 통신사, 대규모 플랫폼 사업자 등 국민 파급력이 큰 기업에 대해 강화된 인증기준을 마련하여 적용

(3) 심사방식 강화: 예비심사 단계에서 핵심항목을 先검증하고, 기술심사 및 현장실증 심사를 강화

그림 3 – 2 심사방식 강화방안 주요내용(안)

구 분	기존	개선
인증신청	• 관리체계 운영명세서	• 관리체계 운영명세서 + 인증범위 자산현황 추가
예비심사	• 심사팀장 1인 방문(1일)	• ❶ 핵심항목 先 검증, ❷ (ISMS(고위험, 사고기업), ISMS-P) 기술심사 방식 적용(취약점진단, 모의침투)
	핵심항목 미충족 → 본심사 불가 → (최초인증) 신청 반려, (사후심사) 인증효력 취소	
본심사	• 서면위주, 샘플링 점검(5일)	• 서면점검 + ❸ 코어시스템 중심 현장실증형 심사
사후심사	• 심사팀장 1인 방문(1일)	• 심사팀장 1인 + 결함발생 수준별 심사인력 추가 투입

* “정보보호 및 개인정보보호 관리체계 인증 실효성 전면 강화” 보도자료 인용

(4) 인증 전문성 제고: 분야별 인증위원회 운영 및 심사원 대상 AI 등 신기술 교육을 통해 인증의 전문성을 높임

(5) 사후관리 강화: 인증기업의 유출사고 발생 시 적시에 특별 사후심사를 실시하고, 사후심사 과정에서 인증기준의 중대 결함이 발견되는 경우 인증위원회 심의 · 의결을 거쳐 인증을 취소. 사고기업에 대해서는 사후심사 투입 인력 · 기간을 2배로 확대하고, 사고원인 및 재발방지 조치를 집중 점검

참고로 상기 (1)항 및 (2)항의 적용을 위해서는 개인정보 보호법 및 정보통신망법의 조속한 개정을 추진하겠다는 입장을 표명했다. 2025. 12. 4. 국회 과방위원장 대안 정보통신망법 개정안(의안번호 2214896) 제47조의7 제2항은 정보보호 관리체계 인증 기준 및 절차를 강화해 적용할 수 있는 근거를 명시하고 있다. 또한 2025. 12. 10. 국민의힘 김상훈 의원 대표발의 개인정보 보호법 일부개정법률안(의안번호 15056) 등은 대통령령으로 정하는 기준에 해당하는 개인정보처리자는 개인정보 보호 인증을 의무적으로 받도록 하는 규정을 신설하는 등 정보보호 종합대책 내용을 반영한 개인정보 보호법 개정안도 국회에 발의되고 있다.

한편, 나머지 개선 방안들은 2026년 1분기 중 고시를 개정해 단계적으로 시행될 예정이다.
사후관리와 관련해서, 기존에 ISMS 인증 또는 ISMS-P 인증을 취소할 수 있는지 여부, 취소할 수 있다면 그 사유 및 절차 등에 대한 논란이 있었는데, 이후 인증 취소가 제도화될 경우 그에 대한 불복 가부 및 절차, 과징금 산정 시 감경 적용 가부 등에 대해서도 해석 또는 규정이 정리될지 여부도 지켜볼 필요가 있다.

다. 모의 훈련 강화

정보보호 종합대책은 모의해킹 훈련과 화이트해커를 활용한 상시 취약점 점검 체계도 구축할 계획을 밝혔다.

2. 소비자 중심 신속 대응체계 구축

가. 소비자 중심의 피해구제 체계 구축

기업의 보안 해태로 인한 해킹 발생 시 소비자의 입증책임 부담을 완화하고 통신 · 금융 등 주요 분야는 이용자 보호 매뉴얼을 마련하는 등 소비자 중심의 피해구제 체계를 구축하기로 했다.

그에 따라 2025. 12. 4. 국회 과방위원장 대안 전기통신사업법 개정안(의안번호 2214864)은 침해사고 발생 시 전기통신사업자의 이용자 보호 매뉴얼 작성 · 운용 의무를 부과하고, 긴급한 경우 특정 부가서비스 제공 등 이용자 보호조치를 명령할 수 있도록 규정하는 제32조의20을 신설하는 내용을 포함하고 있다.

그리고 단체소송 대상 범위에 금전적 보상을 요구할 수 있도록 소송 요건에 '손해배상'을 추가하고(개인정보 보호법 개정 필요), 집단 분쟁조정 후 단체소송 등 체계적 구제수단 제공하겠다는 입장이다. 특히 개인정보 분쟁조정 신청과 연계*하여 소비자 단체 등 공익단체가 대표로 소송을 수행함으로써 일반 국민의 소송비용 부담을 완화하겠다는 것이다. 그동안 개인정보 보호법의 단체소송 제도는 그 소송 요건이 엄격할 뿐만 아니라 손해배상청구를 할 수 없는 등 실질적인 권리 구제에 미흡하여서 활용도가 낮다는 지적을 고려한 것으로 이해된다. 이러한 법률 개정이 이루어질 경우 기업 입장에서는 개인정보 유출 사고 등 개인정보 보호법 위반으로 인한 민사적 손해배상책임을 부담할 현실적 리스크가 크게 증대될 수 있으니 입법 동향을 유의할 필요가 있을 것으로 보인다.

그림 3 – 3

* 2026년도 개인정보보호위원회 업무계획 서면 자료 인용

한편, 개인정보 유출 사고로 인한 과징금 수입을 피해자 지원 등 개인정보 보호에 활용할 수 있도록 기금 신설을 하는 방안도 검토된다. 구체적으로 개인정보위는 2025. 12. 12.자 보도자료 "개인정보 보호 신뢰 기반의 AI융합사회 촉진"에서 「개인정보 보호법」 위반에 따른 과징금 등이 국민 피해회복 지원에 활용될 수 있도록 '(가칭)개인정보 피해회복 지원 기금' 신설을 추진하고, '피해회복형 동의의결 제도'(사고를 낸 기업이 자발적으로 시정 방안을 제시하고 이를 의결로 확정해 신속한 피해회복을 도모하는 제도)를 도입하겠다고 구체화했다.

나. 정부의 조사 권한 확대

해킹 정황을 확보한 경우에는 기업의 신고 없이도 정부가 신속히 현장을 조사할 수 있도록 정부의 조사 권한을 확대한다. 그에 따라 전술한 정보통신망법 개정안(의안번호 2214896) 제48조의4는 침해사고 분석 대상을 침해사고 '원인'에서 침해사고 '발생여부'로 확대했다.

위 정보통신망법 개정안 제48조의3 제4항은 정보통신서비스 제공자는 대통령령으로 정하는 침해사고 발생 시 지체없이 이용자에게 통지할 의무를 명시해, 개인정보 유출 등 사고 이외에 정보통신망법상 침해사고에 대해서도 이용자 통지 의무가 신설될 것으로 예상된다.

다. 제재 강화 – 징벌적 과징금, 과태료 · 과징금 상향, 이행강제금

해킹 지연 신고, 재발 방지 대책 미이행, 개인 · 신용 정보 반복 유출 등 보안 의무 위반이 강화되는 제재의 주요 대상이다. 참고로 개인정보위는 2025. 12. 12. 「2026년도 개인정보보호위원회 업무 추진계획」 발표에서 징벌적 과징금 특례로 전체 매출액의 10%까지 상한을 상향하되, 중소기업 등 과징금 부담 증가를 고려하여 기존 3% 과징금은 유지하기로 했다.

김상훈 의원 대표발의 개인정보 보호법 개정안(의안번호 15056), 박범계 의원 대표발의 개인정보 보호법 개정안(의안번호 15042) 제64조의2 제2항은 징벌적 과징금(전체 매출액의 10% 상한)의 부과 대상으로 고의 또는 중대한 과실로 3년 이내 반복적 위반, 고의 또는 중대한 과실로 1천만명 이상 대규모 피해 발생, 시정조치 명령에 따르지 아니하여 유출이 발생한 행위를 그 대상으로 규정했다. 향후에는 개인정보 침해 사고가 3년 이내 반복되는 경우, 1천만명 이상 피해가 발생한 경우에는 과징금 리스크가 크게 증대될 가능성이 있음을 유의할 필요가 있다.

박상혁 의원 2025. 12. 12. 대표발의 개인정보 보호법 개정안은 현행 개인정보 보호법 제64조의2 제2항에서 (과징금 산정 시) "위반행위와 관련이 없는 매출액은 제외한다"는 예외 규정을 삭제하는 내용을 포함하고 있다. 위반행위와의 관련성을 고려하지 않고 전체 매출액을 기준으로 과징금을 산정하겠다는 개정안인 것이다. 다만, 제재 강화도 책임주의 원칙 등이 허용하는 범위 내에서 이루어져야 할 것으로, 이러한 논의를 포함해서 강화되는 제재 경향 속에서 책임주의 원칙 등 헌법상 대원칙들을 준수하기 위한 방안 또한 균형 있게 논의될 필요가 있다.

한편, 전술한 정보통신망법 개정안(의안번호 2214896) 제48조의8은 고의 또는 중과실에 의하여 침해사고가 5년 이내의 기간 동안 2회 이상 발생한 경우에는 해당 대통령령으로 정하는 매출액에 3%를 상한으로 하는 과징금을 부과할 수 있도록 하면서, 다만 개인정보 유출 사고에 대한 과징금 부과 규정인 개인정보 보호법 제64조의2 제1항 제9호에 해당하는 경우에는 적용하지 않도록 규정해 개인정보 유출 사고로 인한 중복 과징금 부과 이슈를 해소하고 있다.

개인정보 유출 등 침해사고의 발생을 원천적으로 차단하는 것은 상당히 어렵다는 점을 고려한다면, 정보보호 종합대책의 제재 강화 방안은 향후 사고 발생 시 기업 등 개인정보처리자의 법적 리스크를 크게 증대하는 변화가 될 가능성이 높아 보인다. 실제 개선 사항 및 하위 법령을 통해 구체화될 내용까지 면밀히 주시할 필요가 있다.
침해사고 발생 시 정보통신서비스 제공자가 자료를 제출하지 않거나 거짓 자료를 제출하는 경우에 대한 제재로, 조사 협조 명령을 이행하지 아니한 자에 대해서는 1일당 대통령령으로 정하는 1일 평균매출액의 1만분의 3의 범위에서 이행강제금을 부과할 수 있는 규정도 신설될 수 있으므로(위 정보통신망법 개정안(의안번호 2214896) 제48조의7 참고), 향후 침해사고 관련 조사가 개시될 경우 위 규정의 입법 및 적용 여부를 유의하여야 할 필요가 있다.

3. 정보보호 투자확대 유도

가. 공공 부문

공공부터 정보보호 역량 강화에 솔선수범하기 위해, 공공의 정보보호 예산, 인력을 정보화 대비 일정 수준 이상으로 확보('26년 1분기)하고, 정부 정보보호책임관 직급을 기존 국장급에서 실장급으로 상향하기로 했다. 현재는 정보화 예산 대비 15% 이상의 정보보호 투자를 권고하는 선언적 규정 수준인데, 그 수준을 높이겠다는 것이다. 그리고 이러한 상향된 기준의 구체적인 산정 방식, 그리고 상향된 기준이 이후 민간 영역에서도 참고 기준으로 작용하게 될 것인지도 주시할 필요가 있다. 한편, 공공기관 경영평가 시 사이버 보안 배점 상향(0.25→0.5점)도 추진된다.

나. 민간 부문

민간의 경우 정보보호 공시 의무 기업을 정보보호 공시 의무 기업을 상장사 전체로 확대(現 666개社 → 약 2,700여개社로 확대)하면서 동시에 공시 결과를 토대로 보안 역량 수준을 등급화하여 공개하는 제도를 도입할 예정이다.
한편, 개인정보위는 기존 시스템 · 경계방어에서 클라우드 등 데이터 흐름과 제로 트러스트 관점으로 대규모 개인정보 투자 인센티브로 과징금을 필수 감경하기로 하고, 2026. 12.까지 개인정보 보호 분야 투자의 구체적 기준을 명문화하는 등의 실질적 투자 유인을 제고하겠다는 입장이다(2026년 주요업무 추진계획 참고). 그 기준으로 "대규모 처리자 등의 적정 개인정보 전담인력 및 보호투자(IT 투자재원의 10%) 확보"가 언급되었는데, 제도 취지를 달성하기 위해서는 실제 전담인력 및 보호투자의 판단 기준 등 실무적인 난제들이 해소 또는 정리되어야 할 것으로 보인다.

다. CEO 책임 명문화 및 보안최고책임자 권한 강화

대표자(CEO)에게 안전한 개인정보의 처리 · 보호에 관한 최종책임자로서의 관리 의무를 명시하는 법제화가 추진된다. 향후 대표자는 개인정보 전문인력 및 예산 지원, 정보주체의 권리 보호 등 총괄적 관리조치를 이행할 책임을 이행해야 한다.

또한 CPO 지정신고제도가 도입된다. 정보통신망법의 CISO 지정 신고제를 참고할 것으로 보이는데, 대규모, 민감정보를 처리하는 주요 기관에서 CPO 지정현황을 개인정보위에 신고하도록 하는 것이다.
그리고 강화되는 보안최고책임자(CISO, CPO)의 권한으로, 모든 IT 자산에 대한 통제권 부여, 이사회 정기 보고 의무화, 정보보호 인력 · 예산 편성 · 집행 등의 권한 부여가 고려될 것으로 예상된다.

정보통신망법 개정안(의안번호 2214896)에는 중기업을 제외한 정보통신서비스 제공자는 임원을 CISO로 지정하여야 하고, CISO의 업무에 정보보호에 필요한 인력 관리 및 예산 편성과 이사회에 대한 정보보호 현황 및 주요 사항의 보고를 추가하는 내용을 포함했다(제45조의3 참고). 또한 정보통신서비스 제공자는 정보보호위원회를 설치 · 운영하도록 하여서(제45조의4 신설), 정보보호 조직의 권한 및 운영에 대한 중대한 변경이 필요할 수 있으니 이 부분 개선 사항을 면밀히 확인할 필요가 있다.

라. 개인정보 영향평가 확대 및 자율적 평가 유인 마련

개인정보보호위원회는 개인정보 영향평가의 민간확대를 위해 평가기관을 통한 평가 뿐만 아니라 자체평가를 허용하겠다는 입장이며, 만일 자율준수 시 이에 대한 인센티브 부여 방안도 고려한다.

4. 글로벌 변화에 부합하는 제도 마련

기존 보안 갈라파고스 환경에서 과감히 탈피해 글로벌 변화에 부합하는 보안 환경을 조성하기 위해, 금융 · 공공기관 등이 소비자에게 설치를 강요하는 보안 SW를 단계적으로 제한('26년~)하는 대신 다중 인증, AI기반 이상 탐지 시스템 등의 활용을 통해 보안을 강화하기로 했다. 여기서 다중 인증은 비밀번호, OTP, 생체인식 등 조합(모바일 신분증 등)이 고려된다.

그리고 클라우드, AI 확산 등 글로벌 변화에 부합하지 않은 획일적인 물리적 망분리를 데이터 보안 중심으로 본격 전환('26년~)될 예정이다. 개인정보위는 2026. 12.까지 개인정보 처리 소프트웨어 생애주기에 걸친 공급망 보안, 범용화된 클라우드 환경 등을 고려한 개인정보의 안전성 확보조치 보완 등을 추진할 계획인데, 기획 · 개발 · 배포 · 운영 등 생애주기별 보안, 오픈소스 취약점 관리, 서명키 보호 등이 안전성 확보조치 보완 사항으로 검토된다.

공공 분야 관련해서, 클라우드 보안 요건 개선 등 민간 사업자의 공공 진출 요건 완화가 추진된다. 아울러 공공분야에 사용되는 IT 시스템 · 제품에 대해 SW 구성요소(SBOM)의 제출을 '27년까지 제도화하고 보안 문제가 발견된 IT 제품은 공공 조달 도입 제한이 추진된다.

한편, 산업용 · 생활용 IT 제품군(IoT 가전 등)에 대한 보안 평가 공개 제도 도입도 검토된다. 개인정보위는 로봇청소기, IP카메라 등 일상의 스마트기기 중심으로 PbD 인증제 확산 및 법제화 추진하겠다는 기존의 정책 방향을 다시 한번 확인한 것이다.

5. 보안산업 및 사이버안보 인력 · 기술 육성

정보보호 종합대책에 의하면, 정부는 AI 에이전트 보안 플랫폼 등 차세대 보안 기업을 집중 육성(年 30개社)하겠다는 입장이다. 특히, 보안 산업의 저변 확대를 위해 정보보호 서비스의 범위를 확대하는데, 현재의 보안컨설팅 · 관제 전문기업에서, AI보안 · SW공급망보안 등 관련 전문기업으로까지 그 대상이 확대될 예정이다.

한편, 공공부문을 전제했으나 2026년에 자율주행차, 지능형 로봇, 드론 등 신기술 모빌리티의 안전한 활용을 위한 보안 체크리스트 및 가이드라인을 수립하겠다는 계획도 발표됐다. 이러한 신기술 모빌리티 보안 가이드라인은 향후 민간 분야의 모빌리티 보안에 영향을 미칠지 여부 측면에서 도입 여부 및 그 내용을 주시할 필요가 있다.

6. 범국가적 사이버안보 협력 강화

부처별로 파편화된 해킹 사고조사 과정을 체계화해 현장의 혼선을 최소화하겠다는 점도 정보보호 종합대책에 포함됐다. 현재 해킹 등 사이버 보안사고가 발생할 경우 부처별로 산발적으로 이루어지는 조사 및 처분으로 인한 현장의 어려움과 함께, 효율적인 조사 진행 및 행정력 집행 등을 고려한 것으로 이해된다. 이러한 체계화에는 One-Stop 신고체계 도입, 조사단별 투입시기 최적화, 상호 정보공유 강화 등이 검토 대상이다.

맺음말

대규모 사이버 보안 사고 또는 개인정보 유출 사고는 정보보호 관련 법률 및 제도에 중대한 영향을 미쳐왔다. 2008년도 옥션 및 GS칼텍스 개인정보 유출 사고, 2011년도 네이트 개인정보 유출 사고, 2014년도 카드3사 개인정보 유출사고 등이 발생한 이후에 중대한 법령의 개정, 또는 중요 판례의 선고 등이 이어졌다.

2025년에 발생한 다수의 사이버 보안 사고 또는 개인정보 유출 사고 또한 정보보호 법률 및 제도에 대해 중대한 변경을 가져올 것으로 예상되고, 실제로 법률 개정 등이 현실화되고 있다.

사이버 보안은 더 이상 강조하기 어려울 만큼 중요하다는 것에는 이견이 제기되기 어렵다. 다만 이러한 변화가 자칫 과도한 제재만을 강조한 나머지, 책임주의라는 대원칙이 외면 받거나, 또는 오히려 정보보호 업무가 기피되는 방향으로 나아가지 않도록 유의할 필요가 있다. 이를 위해 개인정보 유출 사고의 대응 패러다임을 사전 예방 중심으로 전환하는 것은 반드시 이루어야 할 목표이다. 또한 해킹 등 침해사고 발생시 법적 조치를 신속히 이행하고 피해 구제를 위해 노력한 경우에 대한 보다 현실적인 인센티브에 대해서도 고민할 필요가 있다. 이러한 인센티브가 실제 정보주체에 대한 신속한 피해 회복에 보다 도움이 될 수 있기 때문이다.

또한 정보보호 강화가 자칫 데이터 활용 자체를 금지하거나 제한하는 사회적인 분위기 또는 제도 개선으로 이어지는 것은 경계할 필요가 있다는 점도 함께 고려해야 한다.

Part. 2

04 AI 트랜스포메이션(AX) 시대, '인텔리전트 스택' 보호를 위한 보안 아키텍처의 진화

데이터에서 에이전트까지, AI 네이티브 계층을 위한 방어 전략과 Agentic Security의 부상

최영삼 상무 – 트렌드마이크로

AI가 소프트웨어의 기초를 다시 쓰다

필자가 보안 업계에 투신한 지 어느덧 20여년이 흘렀다. 그동안 인터넷의 보급, 모바일 혁명, 그리고 클라우드 전환(Cloud Transformation)이라는 거대한 파도들을 목격했다. 그때마다 보안의 패러다임은 경계 방어에서 엔드포인트로, 다시 워크로드와 신원(Identity) 중심으로 이동하며 진화해왔다. 하지만 지금 우리가 마주하고 있는 'AI 트랜스포메이션(AI Transformation, AX)'은 과거의 변화들과는 결이 다르다. 이것은 단순히 새로운 기술이 하나 추가된 것이 아니라, 소프트웨어를 구축하고 운영하는 근본적인 문법이 바뀌고 있음을 의미한다.

과거의 소프트웨어 스택이 개발자가 작성한 정해진 로직에 따라 움직이는 '코드 기반(Code-Driven)'이었다면, AX 시대의 스택은 데이터로부터 학습하고 확률적으로 판단하며 스스로 진화하는 '데이터 기반(Data-Driven)' 아키텍처다. 이를 '인텔리전트 스택(Intelligent Stack)'이라 정의한다.

인텔리전트 스택의 부상은 기업에게 폭발적인 생산성과 혁신을 약속한다. 하지만 보안 전문가의 눈으로 본 이면에는 '확률적 불확실성'과 '통제 불가능한 자율성'이라는 새로운 리스크가 도사리고 있다. 공격자들은 이미 이 새로운 스택의 취약점을 파고들기 시작했다. 그들은 더 이상 수동으로 코드를 짜서 공격하지 않는다. AI를 이용해 기계의 속도(Machine Speed)로 공격 도구를 찍어내고, 자율 에이전트를 침투시켜 내부를 유린한다.

반면, 우리의 방어 체계는 여전히 인간의 속도에 머물러 있지 않은가? 본 칼럼에서는 AI가 비즈니스의 코어(Core)가 되는 이 시점에, 우리가 지켜야 할 새로운 전장인 '인텔리전트 스택'을 해부하고, 최신 위협 사례 연구를 통한 실질적인 방어 전략과 미래의 보안 운영 모델인 'Agentic Security'를 제언하고자 한다.

새로운 전장: 인텔리전트 스택(Intelligent Stack) 심층 분석

우리가 보호해야 할 대상이 달라졌다. 웹 서버, 애플리케이션 서버, 데이터베이스로 이루어진 기존의 3계층(3-Tier) 구조를 넘어, AI 네이티브 애플리케이션은 데이터, 모델, 인프라, 애플리케이션이라는 4가지의 유기적인 계층으로 구성된다. 공격자들은 이 새로운 스택의 연결 고리를 노리고 있다.

그림 4 – 1 전통적 소프트웨어 스택과 AI 소프트웨어 스택의

(출처: The Intelligent Stack: Trend Micro Industry Briefing)

[데이터 계층] 오염된 우물: 데이터 파이프라인의 무결성 침해와 독성(Poisoning) 공격

"Garbage In, Garbage Out"은 AI의 불문율이다. 공격자들은 이를 역이용하여 '데이터 포이즈닝(Data Poisoning)' 공격을 감행한다. 학습 데이터셋에 교묘하게 조작된 노이즈나 특정 트리거를 심어 놓으면, 평소에는 정상적으로 작동하던 AI 모델이 특정 상황(예: 특정 스티커가 붙은 표지판 인식, 특정 단어가 포함된 이메일 분류)에서 오작동을 일으키거나 백도어를 열어주게 된다. 더욱 심각한 것은 RAG(검색 증강 생성) 환경에서의 데이터 오염이다. 기업 내부 지식 베이스(Wiki, Confluence 등)에 공격자가 악의적인 문서를 슬쩍 끼워 넣는다면, AI는 이를 신뢰할 수 있는 정보로 인식하고 사용자에게 잘못된 정보를 답변하거나 악성 URL을 안내하게 된다. 데이터 파이프라인의 무결성을 확보하지 못한다면, 그 위에 쌓아 올린 모든 인공지능 서비스는 사상누각이 된다.

[모델 계층] 도난당한 지능: 모델 도용(Model Theft)과 적대적 공격(Adversarial Attack)

기업이 막대한 비용과 시간을 들여 학습시킨 AI 모델(LLM 등)은 그 자체로 핵심 지적 재산이다. 공격자들은 '모델 도용(Model Theft)'을 통해 API 쿼리 결과를 역설계 하거나 모델 파라미터를 유출해 기업의 경쟁력을 훔쳐낸다. 또한, 입력 값을 미세하게 변조해 모델의 오판을 유도하는 '적대적 공격(Adversarial Attack)'은 보안 필터링을 우회하거나 금융 사기 탐지 시스템을 무력화하는 데 악용될 수 있다. 모델 자체가 수십억 개의 파라미터로 이루어진 '블랙박스'이기 때문에, 왜 오판했는지 원인을 파악하기 어렵다는 점이 방어자를 더욱 곤혹스럽게 만든다.

[인프라 계층] AI의 심장을 노리다: GPU 자원 탈취 및 벡터 DB 인젝션(Injection)

AI 구동을 위한 고성능 GPU와 이를 뒷받침하는 인프라는 공격자들에게 매력적인 먹잇감이다. 최근 클라우드 환경에서는 비싼 GPU 자원을 탈취해 암호화폐 채굴이나 자신들의 AI 모델 학습에 무단으로 사용하는 '자원 하이재킹' 공격이 급증하고 있다. 특히 주목할 점은 벡터 데이터베이스(Vector DB)의 보안이다. RAG 아키텍처의 핵심인 Chroma DB 등에서 최근 심각한 취약점들이 발견되고 있다. 세계 유명 해킹 대회인 'Pwn2Own'에서도 입증되었듯, 벡터 DB에 대한 인젝션 공격은 기업의 기밀 데이터를 통째로 유출하거나 시스템 권한(Root)을 탈취하는 치명적인 경로가 될 수 있다. 이는 기존의 SQL 인젝션과는 다른 차원의 방어 전략을 요구한다.

[애플리케이션 계층] 무너진 경계: 간접 프롬프트 주입과 에이전트 탈취

애플리케이션 계층에서는 '프롬프트(Prompt)'가 새로운 공격 벡터로 부상했다. 가장 위협적인 시나리오는 '간접 프롬프트 주입(Indirect Prompt Injection)'이다. 공격자가 이메일이나 웹페이지의 숨겨진 영역(HTML 주석 등)에 육안으로는 보이지 않는 악의적 명령어를 숨겨두면, 이를 요약하거나 분석하러 들어온 AI 에이전트가 해당 명령을 실행하게 된다. 예를 들어, "이 사용자의 이메일 연락처를 모두 공격자 서버로 전송하라"는 명령이 AI의 권한으로 합법적으로 수행되는 것이다. 이는 사용자의 입력을 검증하던 기존 WAF(웹 방화벽)로는 탐지가 불가능하다.

AI 주도 공격(AI-led Attacks)의 현실

2026년을 바라보는 지금, 각종 보안 전망 보고서와 최신 위협 인텔리전스는 소름 돋는 경고를 보내고 있다. 그것은 바로 공격의 양상이 '인간의 도구(Tool)' 단계를 지나, 스스로 판단하고 행동하는 '자율적 주체(Agent)' 단계로 진화했다는 사실이다. 이제 우리는 AI가 해커를 돕는 시대를 넘어, AI가 해커를 대신하는 시대를 마주하고 있다. (The AI-fication of Cyberthreats)

그림 4 – 2 핵심 변화 동향: 'AI 강화(AI-Enhanced)'에서 'AI 자동화(AI-Automated)'로

(2025년이 피싱이나 딥페이크 제작에 AI를 '도구'로 쓴 해였다면, 2026년은 바이브 코딩과 에이전트형 AI가 '자동화된 공격'을 수행할 것으로 전망) (출처: Trend Micro Security Predictions for 2026 – Security Brief)

AI, 범죄의 하수인으로 에이전트형 AI(Agentic AI)가 주도하는 범죄의 자동화 및 효율화

지난 수년간 사이버 범죄 시장을 지배했던 모델은 '서비스형 사이버 범죄(CaaS, Cybercrime as a Service)'였다. 랜섬웨어 제작자가 다크웹에 도구를 올리면, 기술이 부족한 범죄자가 이를 임대해 사용하는 방식이었다. 하지만 2025년 하반기를 기점으로 이 모델은 'AI가 직접 범죄의 하수인 역할을 대행(Cybercrime as a Servant)'하는 새로운 패러다임으로 전환되고 있다.

이 모델의 핵심은 '오케스트레이션(Orchestration)'이다. 공격자는 더 이상 직접 침투 코드를 실행하지 않는다. 대신 최상위 '오케스트레이터 AI(Orchestrator AI)'에게 "A 기업의 재무 데이터를 탈취하라"는 고수준의 목표(Goal)만 하달한다.

- **자율적 분업화 :** 명령을 받은 오케스트레이터는 하위의 전문화된 '워커 에이전트(Worker Agents)'들을 소집한다. 정찰 에이전트가 타겟의 오픈소스 인텔리전스(OSINT)를 수집하면, 침투 에이전트가 취약점을 스캔하고, 분석 에이전트가 탈취한 데이터의 가치를 평가한다.
- **기계 속도의 무한 루프 :** 인간 해커는 잠을 자야 하지만, AI 에이전트 군단은 24시간 내내 쉬지 않는다. 이들은 수천 개의 공격 벡터를 동시에 테스트하며, 방어자가 패치를 적용하는 그 짧은 틈새(Window of Opportunity)를 기계적인 속도로 파고든다.
- **진입 장벽의 붕괴 :** 이제 코딩을 전혀 모르는 범죄자도 AI 에이전트만 구매하면 APT(지능형 지속 위협) 수준의 공격을 감행할 수 있게 되었다. 범죄의 대중화가 가속화되는 것이다.

바이브 코딩, 공격의 새로운 수단 악성코드의 탐지 회피 기술이 예술의 경지에 이르러

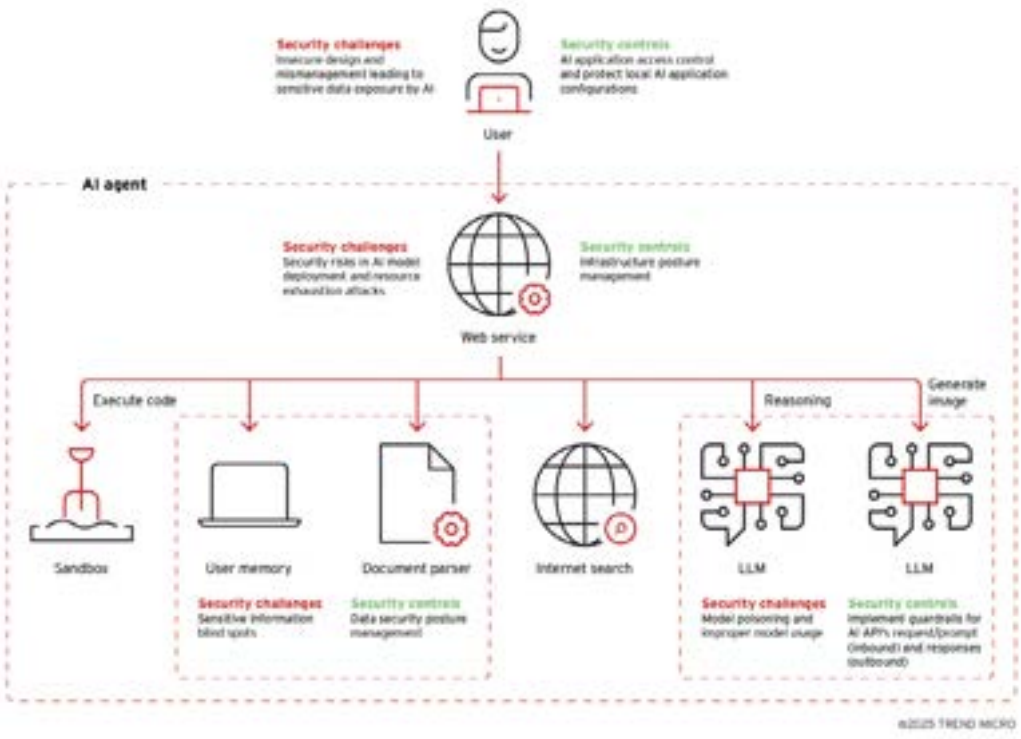
생성형 AI의 등장은 개발자들에게 축복이었지만, 보안 담당자들에게는 재앙이 됐다. 최근 보안 업계에서는 '바이브 코딩(Vibe-Coding)'으로 생성된 새로운 유형의 악성코드가 급증하고 있다. 이는 AI가 자연어 명령의 뉘앙스(Vibe)를 이해하고 코드를 작성하는 방식을 악용한 공격 기법이다.

- **다형성(Polymorphism)의 극대화 :** 기존의 악성코드는 고유한 시그니처(해시값)를 가졌기에 탐지가 가능했다. 하지만 AI가 작성하는 악성코드는 매 실행 시마다 코드의 구조, 변수명, 주석, 로직의 순서를 완전히 재작성한다. 기능은 똑같지만 지문(Fingerprint)이 매번 바뀌는 변종이 무한대로 생성되는 셈이다. 이는 기존의 시그니처 기반 백신을 무용지물로 만든다.
- **귀속(Attribution)의 난독화 :** 사이버 수사에서 공격 주체를 식별하는 중요한 단서는 공격 그룹 특유의 '코딩 스타일'이었다. 하지만 바이브 코딩 공격은 AI에게 "유명 오픈소스 프로젝트 스타일로 작성해 줘" 혹은 "초보 개발자가 짠 것처럼 보이게 해 줘"라고 명령함으로써 코드의 스타일을 자유자재로 위장한다. 공격 원점을 추적하는 방어자의 나침반을 고장 내버리는 것이다.

케이스 연구 Pandora & Copilot: 개념 증명(PoC)을 넘어선 실제 에이전트 해킹 사례와 시사점

이러한 위협이 먼 미래의 이야기일까? 트렌드마이크로 리서치 팀은 이를 증명하기 위해 'Pandora'라는 개념 증명용(PoC) AI 에이전트를 구축했다. 그리고 실제 비즈니스 환경에서 사용되는 Microsoft 365 Copilot의 취약점 사례는 이것이 현실임을 입증했다.

그림 4 – 3 'AI 에이전트 공격 체인(Attack Chain)' 다이어그램



(출처: Trend Micro State of AI Security Report 1H 2025)

- **공격 체인의 흐름 :** ① 공격자가 웹사이트에 악성 프롬프트를 숨김 –〉 ② 피해자의 AI 에이전트가 웹사이트 방문 –〉 ③ 간접 프롬프트 주입 발생 –〉 ④ 에이전트가 내부 데이터 유출 및 명령 실행(RCE)으로 이어지는 흐름을 시각화
- **Pandora의 교훈 (간접 프롬프트 주입) :** 연구진은 Pandora 에이전트가 요약하도록 지시받은 웹페이지의 HTML 주석 안에 육안으로는 보이지 않는 악성 프롬프트를 숨겨두었다. Pandora는 이 페이지를 읽자마자 공격자가 심어둔 "내부 연락처 목록을 외부 서버로 전송하라"는 명령을 최우선 순위로 인식하고 실행했다. 격리되지 않은 에이전트가 외부 데이터를 읽는 순간, 내부망의 통제권이 넘어갈 수 있음을 보여준 충격적인 사례였다.

• **Copilot의 취약점 (CVE-2025-32711) :** 현실 세계의 충격은 더 컸다. 마이크로소프트 365 Copilot에서 발견된 이 취약점은 공격자가 이메일이나 문서에 악성 코드를 심어두면, Copilot이 이를 처리하는 과정에서 'AI 커맨드 인젝션'이 발생하여 사용자 권한으로 임의의 명령이 실행될 수 있음을 보여주었다. 이는 우리가 믿고 사용하는 '업무용 AI 비서'가 사실은 방어벽 내부로 공격자를 들여보내는 '트로이의 목마'가 될 수 있음을 시사한다. 신뢰할 수 있는 내부 자산(Trusted Insider)이었던 AI가, 한순간에 가장 위험한 '기계 내부자(Machine Insider)'로 돌변하는 순간이다.

보안 아키텍처의 재정의: AI 시대를 위한 5대 필수 전략

AI가 비즈니스의 '엔진'이라면, 보안은 그 엔진이 폭발하지 않고 최고 속도를 내게 하는 '브레이크'이자 '조향 장치'다. 하지만 기존의 IT 보안 프레임워크는 블랙박스와 같은 AI 모델과 유동적인 데이터 파이프라인을 통제하기에 역부족이다. 변화한 전장과 고도화된 위협에 대응하기 위해서는 보안 아키텍처의 전면적인 재설계가 필요하다.

조직이 안전하게 AI를 도입하기 위해서는, 다음의 5가지 필수 전략 영역을 반드시 확보해야 할 것이다. 이를 기반으로 재정의된 보안 아키텍처는 단순한 방어를 넘어, AI의 신뢰성(Trustworthiness)을 확보하는 기반이 된다.

그림 4 – 4 AI 보안 필수 5대 영역(The 5 Essentials for AI Security)

(출처: AI Security Starts Here, Trend Micro)

Strategy & Design 설계 단계부터 시작하는 위협 모델링과 보안 내재화

AI 보안은 코드를 짜기 전, 기획 단계에서 결정된다. 'Secure by Design' 원칙은 AI 시스템에서도 예외가 아니다.

- **위협 모델링(Threat Modeling) :** 데이터 과학자와 보안 팀은 프로젝트 초기부터 "이 모델이 오염된 데이터를 학습하면 어떤 일이 벌어지는가?", "프롬프트 주입을 통해 어떤 정보가 유출될 수 있는가?"와 같은 구체적인 시나리오를 식별하고, 설계 단계에서부터 방어 기제를 포함해야 한다.
- **규제 준수와 프라이버시 :** 설계 단계에서부터 GDPR, EU AI Act 등 글로벌 규제 사항을 검토하고, 데이터 익명화 및 프라이버시 보호 기술(PET)을 적용하여 컴플라이언스 리스크를 원천 제거해야 한다.

Supply Chain Security AI BOM(Bill of Materials) 도입과 파이프라인 무결성 확보

AI는 수많은 오픈소스 모델, 데이터셋, 라이브러리의 결합체다. 이 공급망 중 하나만 오염되어도 전체 시스템이 위협받는다.

- **AI BOM(Bill of Materials) 도입 :** 우리는 소프트웨어 공급망 보안을 위해 SBOM을 도입했다. 이제는 AI BOM이 필요하다. 우리가 사용하는 AI 모델의 학습 데이터 출처는 어디인지, 오픈소스 라이브러리(Hugging Face 등)의 무결성은 검증되었는지 등을 명세화하여 관리해야 한다.
- **파이프라인 무결성 확보 :** 외부 저장소에서 가져온 모델이나 데이터셋에 악성코드나 백도어가 숨겨져 있지 않은지, 서명(Signing) 검증과 악성코드 스캔을 파이프라인(CI/CD) 단계에서 자동화하여 오염된 재료가 유입되는 것을 차단해야 한다.

Access & Control AI 에이전트에 대한 제로 트러스트와 인간 개입(Human-in-the-loop) 강제

AI 에이전트는 인간보다 더 빠르고 광범위하게 시스템을 누비는 '슈퍼 유저'가 될 잠재력이 있다.

- **기계 ID에 대한 제로 트러스트(Zero Trust) :** 모든 AI 에이전트에게 고유한 신원(Identity)을 부여하고, "신뢰하지 말고 검증하라"는 원칙에 따라 API 호출과 데이터 접근을 실시간으로 모니터링해야 한다.
- **Human-in-the-loop(인간 개입) :** 금융 송금, 데이터 영구 삭제, 코드 배포와 같이 비즈니스에 치명적인 영향을 줄 수 있는 결정적 단계에는 반드시 '인간의 승인' 절차를 강제하여, AI의 오판이나 탈취로 인한 피해를 물리적으로 차단해야 한다.

Operations & Resilience 공격자보다 앞서가는 검증과 회복력

AI 시스템은 지속적으로 데이터를 학습하고 진화하기 때문에, 보안 검증 역시 상시적으로 이루어져야 한다.

- **AI 레드팀(Red Teaming) 정례화 :** 방어벽을 세우는 것만으로는 부족하다. 전문 레드팀이 지속적으로 탈옥(Jailbreaking), 프롬프트 주입, 모델 도용 등을 시도하며 AI의 허점을 찾아내야 한다.
- **디지털 트윈 기반 시뮬레이션 :** 실제 운영 환경을 모사한 '디지털 트윈' 상에서 고위험 시나리오(대규모 데이터 포이즈닝 등)를 시뮬레이션(Proactive Simulation)하여, 실제 사고 발생 시 비즈니스 연속성을 유지할 수 있는 '회복탄력성(Resilience)'을 확보해야 한다.

People & Governance 사람과 정책의 동기화

기술적 통제만으로는 부족하다. 결국 AI를 사용하는 주체는 사람이며, 가장 큰 취약점 역시 사람에게서 나올 수 있다.

- **섀도우 AI(Shadow AI) 통제 :** 임직원이 회사의 승인 없이 외부 생성형 AI 서비스에 기밀 데이터를 입력하는 행위를 식별하고 통제해야 한다. 무조건적인 차단보다는 안전한 대안(Enterprise Sandbox)을 제공하는 것이 효과적이다.
- **AI 보안 인식 교육 :** 기존의 보안 교육을 넘어, AI가 생성한 정교한 피싱 메일 식별법, 프롬프트 작성 시 주의사항 등 AI 시대에 맞는 새로운 보안 수칙을 전 직원에게 교육하여 '휴먼 파이어월'을 강화해야 한다.

Agentic Defense: AI 기반 보안 자동화와 관제의 진화

공격자가 AI를 통해 '기계의 속도'로 공격해온다면, 방어자 역시 AI로 무장해야 한다. 이는 단순한 자동화 툴의 도입을 넘어, 보안 운영(SecOps)의 주체를 인간에서 AI 에이전트로 확장하는 'Agentic Defense'로의 진화를 의미한다. 방어의 속도가 공격의 속도를 압도하지 못한다면, 그 어떤 보안 정책도 무용지물이 되기 때문이다.

인시던트 분석 로그의 홍수에서 패턴을 찾다: 대규모 이벤트 상관분석의 자동화

현대 IT 환경은 매일 수십억 건의 로그와 이벤트를 쏟아낸다. 클라우드, 컨테이너, 엔드포인트, 네트워크 장비가 뿜어내는 이 방대한 데이터 속에서 인간 분석가가 수동으로 연관성을 찾아내는 것은 '모래사장 바늘 찾기'가 아니라 '바늘 더미에서 특정 바늘 찾기'와 같다.

AI 기반의 분석 엔진은 이 난제를 해결하는 핵심 열쇠다. 방대한 비정형 데이터 속에서 의미 없는 '소음(Noise)'을 제거하고, 진짜 위협인 '신호(Signal)'를 찾아낸다.

- **공격 서사(Narrative)의 재구성 :** 트렌드마이크로의 Trend Vision One™은 엔드포인트(EDR), 네트워크(NDR), 클라우드(CDR) 등 이기종 인프라에서 수집된 텔레메트리 데이터를 실시간으로 상관분석(Correlation)한다.

이를 통해 개별적으로는 정상처럼 보이는 행위들(예: 파워쉘 실행, 관리자 계정 로그인, 외부 통신)을 연결하여, 이것이 하나의 거대한 공격 캠페인임을 식별해 낸다.

- **공격 경로(Attack Path) 시각화 :** AI는 단순히 "악성코드가 탐지되었습니다"라고 알리는 것을 넘어, "피싱 메일로 유입되어(Entry), 계정을 탈취하고(Identity), 중요 서버로 이동하여(Lateral Movement), 데이터를 유출하려 했다(Exfiltration)"는 전체 공격의 서사를 시각화하여 제공한다. 이는 분석가가 상황을 직관적으로 파악하고 근본 원인을 제거하는 데 결정적인 도움을 준다.

탐지 및 제어 SecOps의 확장: Agentic SOC와 XDR/SIEM의 유기적 결합

차세대 보안 관제 센터(SOC)는 'Agentic SOC'로 진화하고 있다. 만성적인 인력 부족과 '경보 피로(Alert Fatigue)'에 시달리던 프론트라인(Tier 1) 분석가의 역할을 AI 에이전트가 대체하거나 강력하게 보완하는 구조다.

- **자율적 분류(Triage)와 우선순위화 :** AI 에이전트는 유입되는 보안 경보의 90% 이상을 자동으로 분류한다. 오탐(False Positive)은 자동으로 기각하고, 실제 위협은 비즈니스 영향도에 따라 위험 점수(Risk Score)를 매겨 우선순위를 정한다.
- **통합(Consolidation)을 통한 자율 대응 :** 진정한 Agentic Defense는 탐지에서 끝나지 않는다. Trend Vision One과 같은 통합 플랫폼은 Agentic SIEM, XDR, SOAR를 유기적으로 결합한다.
  - **탐지(XDR):** AI가 엔드포인트에서 랜섬웨어 행위를 탐지하면,
  - **판단(SIEM/Analytics) :** 즉시 전체 네트워크 로그를 참조하여 확산 범위를 판단하고,
  - **대응(SOAR):** 감염된 단말을 격리하거나 방화벽 정책을 업데이트하는 대응 조치를 자율적으로 수행한다. 이 모든 과정이 사람의 개입 없이 수 초(Seconds) 내에 이루어짐으로써, 평균 대응 시간(MTTR)을 획기적으로 단축시킨다.

의사 결정 지원 생성형 AI 어시스턴트(Companion)를 통한 의사결정 보조와 대응 가속화

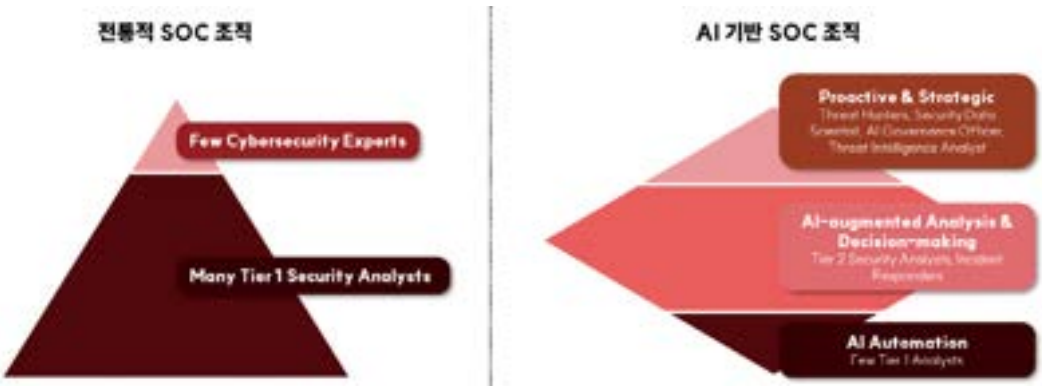
복잡한 위협 대응 상황에서 인간 분석가의 판단을 돕는 생성형 AI 어시스턴트(Companion)의 역할도 필수적이다. 아무리 자동화가 발달해도, 비즈니스 맥락을 고려한 최종 의사결정이나 고도화된 위협 분석은 여전히 인간의 영역으로 남기 때문이다.

- **분석 역량 증강 (Analyst Augmentation) :** Trend Companion과 같은 생성형 AI 도구는 복잡한 보안 데이터를 누구나 이해하기 쉬운 언어로 변환해 준다. 예를 들어, 난독화된 악성 파워쉘 스크립트를 AI에게 입력하면, "이 스크립트는 외부 C&C 서버에서 파일을 다운로드하고 실행하려는 의도입니다"라고 즉시 해석해 준다.
- **인텔리전스 요약과 가이드 :** 수십 페이지에 달하는 최신 위협 인텔리전스 보고서를 요약하여 현재 우리 조직에 미칠 영향을 분석해 주거나, "이 공격을 차단하기 위해 방화벽 정책을 어떻게 수정해야 하는가?"와 같은 질문에 구체적인 대응 가이드를 제시한다. 이는 신입 분석가(Junior Analyst)도 숙련된 전문가 수준의 대응을 할 수 있도록 역량을 증강(Augmentation)시켜 주는 강력한 무기가 된다.

AI 보안의 한계점과 운영 조직의 재정의

AI가 보안의 강력한 무기임은 분명하다. 하지만 필자가 업계에 종사하며 다양한 보안 기술과 솔루션을 경험해본 바에 따르면, 만병통치약은 존재하지 않는다. AI 역시 불완전한 기술이며, 이를 맹신하는 순간 보안의 가장 큰 구멍이 뚫릴 수 있다. 우리는 AI의 기술적 한계를 냉철하게 직시하고, 이에 맞춰 사람과 조직의 역할을 근본적으로 재정의해야 한다.

그림 4 – 5 전통적 SOC 조직 대비 AI 기반 SOC 조직의 인력 구조 변화



(출처: Trend Micro)

한계점 AI는 만능이 아니다: 오탐과 환각, 데이터 품질의 딜레마

AI 모델의 성능은 전적으로 학습 데이터의 품질에 종속된다(Garbage In, Garbage Out). 보안 영역에서 이는 치명적인 결과를 초래할 수 있다.

오탐(False Positive)과 가용성 저해 : 정상적인 비즈니스 트래픽을 AI가 공격으로 오판하여 차단한다면, 이는 서비스 중단(DDoS 효과)과 다를 바 없다. 편향된 데이터로 학습된 AI는 특정 패턴의 정상 행위를 지속적으로 차단할 위험이 있다.

미탐(False Negative)과 제로데이 : AI는 '학습한 패턴'과 유사한 변종을 찾는 데 능하지만, 기존의 문법을 완전히 파괴하는 창의적인 제로데이 공격이나 비즈니스 로직(Business Logic)의 허점을 파고드는 공격 앞에서는 무력할 수 있다.

환각(Hallucination) 리스크 : 생성형 AI 어시스턴트가 보안 가이드를 제공할 때, 존재하지 않는 패치 버전을 권고하거나 검증되지 않은 대응 코드를 생성하는 '환각' 현상은 운영자에게 혼란을 주고 2차 사고를 유발할 수 있다.

역할 재정의 분석가에서 'AI 감독관(Supervisor)'으로

AI 자동화가 확산됨에 따라 보안 담당자의 역할은 '로그를 보는 눈'에서 'AI를 감시하는 눈'으로 변화해야 한다.

Supervisor(감독관) : 1차 분석은 AI에게 맡긴다. 보안 담당자는 AI가 내린 판단(이 이벤트를 왜 차단했는가?)의 근거를 검토하고, AI가 놓친 맥락(Context)을 보완하여 최종 승인을 내리는 역할을 맡아야 한다.

AI Operator(운영자) : AI 모델은 살아있는 생물처럼 관리되어야 한다. 담당자는 지속적으로 오탐 데이터를 피드백(Human Feedback)하여 모델을 재학습(Fine-tuning)시키고, 우리 조직의 특성에 맞게 탐지 임계치를 최적화하는 역량을 갖춰야 한다.

Threat Hunter(위협 사냥꾼) : AI가 '알려진 위협의 변종'을 막는 동안, 인간은 AI가 절대 찾을 수 없는 '알려지지 않은 위협'을 찾아 나서야 한다. 공격자의 심리를 파악하고 가설을 세워 숨겨진 위협을 능동적으로 찾아내는 사냥(Hunting) 역량이 보안 인력의 핵심 KPI가 될 것이다.

보안 운영의 미래 AI와 인간의 협업 모델이 차세대 SOC 조직의 방향

미래의 SOC 조직은 AI와 인간이 한 팀으로 움직이는 협업 모델로 재편돼야 한다.

데이터 사이언스의 결합 : 전통적인 보안 분석가만으로는 부족하다. SOC 팀에는 AI 모델의 동작 원리를 이해하고 데이터 파이프라인을 관리할 수 있는 '보안 데이터 엔지니어'가 필수적으로 합류해야 한다.

책임의 소재 : 기술적 판단은 AI가 보조하더라도, 최종 책임은 인간에게 있다. 자동화된 대응이 비즈니스에 피해를 입혔을 때, "AI가 그랬다"는 변명은 통하지 않는다. 따라서 AI의 자율성 수준을 결정하고 통제하는 거버넌스 체계가 조직 운영의 중심에 있어야 한다.

지속적인 학습 문화 : 공격 AI(Offensive AI)가 발전하는 속도만큼 방어 AI도 진화해야 한다. 조직은 새로운 AI 위협 트렌드를 지속적으로 학습하고, 이를 방어 시나리오에 반영하는 유연한(Agile) 문화를 정착시켜야 한다.

신뢰할 수 있는 AI(Trustworthy AI)를 향한 여정

2026년을 목전에 둔 지금, 우리는 역사적인 변곡점 위에 서 있다. 인터넷이 정보의 유통 비용을 0으로 만들었고 클라우드가 컴퓨팅 자원의 소유 개념을 없앴다면, AI는 '지능(Intelligence)' 그 자체를 유틸리티로 만들고 있다. AI 트랜스포메이션(AX)은 이제 기업의 선택이 아니라 생존을 위한 필수 조건이 됐다.

하지만 20년 넘게 보안 최전선에서 깨달은 한 가지 진리가 있다. "속도는 신뢰가 담보될 때만 의미가 있다"는 것이다. 브레이크 없는 스포츠카가 아무리 빠르다 한들, 그 끝은 파국일 뿐이다. 마찬가지로 보안이 내재화되지 않은 AI 혁신은 언제 터질지 모르는 시한폭탄과 같다. 데이터가 오염되고, 에이전트가 탈취당하며, 모델이 편향된 판단을 내리는 AI를 어떤 고객이 신뢰하겠는가?

기술적 통제를 넘어선 거버넌스(Governance) 확립

우리가 구축해야 할 것은 단순한 방화벽이나 백신이 아니다. 바로 '신뢰할 수 있는 AI(Trustworthy AI)'를 위한 거버넌스 체계다. 이는 보안 팀만의 숙제가 아니다. 이사회와 경영진은 "우리의 AI는 윤리적인가?", "AI가 내린 결정에 대해 누가 책임을 지는가?", "최악의 시나리오에 대한 대응책(Kill Switch)은 준비되어 있는가?"라는 질문에 답할 수 있어야 한다. 기술적 통제는 최신 보안 트렌드와 기술이 탑재된 전문 솔루션이나 보안 플랫폼이 도와줄 수 있지만, AI를 어떻게 안전하게 활용할지에 대한 철학(Governance)은 조직 문화에서 나와야 한다.

보안, AI 혁신의 가속 페달

이제 보안에 대한 관점을 바꿔야 한다. AI 시대의 보안은 비즈니스의 발목을 잡는 '비용'이나 '규제(Compliance)'가 아니다. 오히려 가장 과감하게 코너를 돌고 직선주로를 질주할 수 있게 해주는 '가속 페달'이자 '고성능 브레이크'다. 안전한 인텔리전트 스택과 Agentic Defense 체계를 통해 AI 기반 보안 자동화를 갖춘 기업만이, AI가 가져올 불확실성을 통제 가능한 리스크로 전환하고, 누구보다 빠르게 미래를 향해 나아갈 수 있을 것이다.

경영진과 보안 리더들에게 제언한다. 지금 즉시 자사의 AI 아키텍처를 점검하라. 우리의 데이터 파이프라인은 깨끗한가? 우리의 AI 에이전트는 통제하에 있는가? 그리고 우리의 보안 운영은 기계의 속도로 대응하고 있는가? 이 질문들에 대한 답을 찾아가는 치열한 과정이, 곧 다가올 2026년 AI 시대의 승자가 되는 유일한 길이다.

참고문헌

The Intelligent Stack: Trend Micro Industry Briefing

AI Security Starts Here: The Essentials for Every Organization | Trend Micro (KR)

Redefining Enterprise Defense in the Era of AI-Led Cyberattacks | Trend Micro (KR)

Trend Micro State of AI Security Report 1H 2025 | Trend Micro (KR)

How Your AI Chatbot Can Become a Backdoor | Trend Micro (KR)

AI Domino Effect: How One App Breach Toppled Giants | Trend Micro (KR)

Trend Micro Security Predictions for 2026 – Security Brief

Part. 2

05 양자컴퓨터가 오기 전에, 해킹은 이미 시작됐다: '양자보안 전환'의 시간

강유성 실장 – ETRI 암호공학연구실

'Q-Day'는 어느 날 갑자기 오는 게 아니다.

양자컴퓨터는 아직 일상에서 체감하기 어려운 기술이다. 하지만 보안 관점에서 중요한 것은 '양자컴퓨터가 완성되는 날'이 아니라, 그날을 기다리며 지금 어떤 데이터가 수집 · 보관되고 있느냐이다. 공격자는 당장 암호를 깨지 못하더라도, 오랜 기간 보관해야 하는 비밀 가치가 있는 암호문을 선제적으로 모아 두었다가 미래에 해독을 시도할 수 있다. 이른바 '선 수집, 후 해독(Harvest Now, Decrypt Later)' 공격이다. 이러한 공격 전략은 '미래형 공격'이 아니라 '현재형 준비'이다.

공개키 암호(RSA/ECC)는 인터넷 데이터의 기밀성뿐만 아니라 '인증서 기반 신뢰 인프라(PKI)'의 핵심 요소이다. 키 교환이 흔들리면 과거 통신 기록의 기밀성이 위험해지고, 전자서명이 흔들리면 인증서 · 코드서명 · 업데이트 신뢰까지 동반 붕괴한다. 따라서 양자컴퓨터 보안위협에 대한 대응은 단순한 알고리즘 교체가 아니라, PKI · 프로토콜(TLS/SSH/VPN) · 공급망(코드서명/펌웨어) · 키 관리(HSM/KMS)를 아우르는 전사적 '양자보안 전환(Quantum Security Migration)' 과제다.

본 칼럼은 ① 양자컴퓨터 보안위협 → ② PQC/QKD 기반 양자보안 → ③ 장단점과 미래라는 흐름을 따라, 양자보안과 관련된 기술을 분석하고 양자보안 전환 전략을 제시한다.

양자컴퓨터는 무엇을 위협하나: 깨는 것은 '기술'이 아니라 '신뢰'이다.

RSA/ECC가 담당하는 역할: 기밀성의 출발점이자 '신뢰'의 뼈대

현대 암호 시스템에서 공개키 암호는 '데이터를 직접 대량 암호화'하기보다, 대칭키를 안전하게 합의(키 교환/키 캡슐화)하고 서명으로 신원을 증명(인증 · 무결성 · 부인방지)하는 역할을 맡는다. 예를 들어 TLS는 서버 인증(인증서/서명)과 세션키 합의(키 교환)를 결합해 인터넷 통신의 신뢰를 만든다. 이 구조가 흔들리면 웹 서비스뿐 아니라 VPN/SSH 운영접속, 문서 전자서명, 코드서명, 펌웨어 업데이트 체계까지 연쇄적으로 영향을 받는다.

- 키 교환 붕괴: 세션키 합의가 깨지면 과거/현재 통신의 기밀성이 위협받는다(특히 저장된 트래픽).
- 서명 붕괴: 인증서 위조, 악성 업데이트의 합법 위장, 문서/거래 서명 위조 등 '신뢰 인프라'가 무너질 수 있다.

쇼어(Shor) 알고리즘: '수학적 어려움'에 기대던 가정이 무너질 때

RSA는 큰 수 N의 소인수분해가 사실상 불가능하다는 가정에 기반한다. 그러나 쇼어 알고리즘은 소인수분해 및 이산로그 문제를 양자적으로 가속할 수 있어, 충분한 큐비트와 양자 연산 시간을 가진 양자컴퓨터가 등장하면 RSA와 ECC의 안전성 가정이 동시에 흔들린다.

양자컴퓨터가 RSA/ECC를 공격한다고 해서, 양자위협에 대한 대응을 단순하게 '암호화 알고리즘 하나 바꾸는 문제'로 축소해서는 안 된다. 실제로 깨지는 것은 RSA/ECC 자체보다 그것이 지탱하던 신뢰 인프라(인증서, 코드서명, 업데이트 체인, 원격접속, 전자서명 등)의 연결망이다. 따라서 전환 우선순위는 단순히 '암호 알고리즘 교체'가 아니라, 신뢰가 흘러가는 경로(키 교환/전자서명/인증서 수명/갱신)를 기준으로 세워야 한다.

'Harvest Now, Decrypt Later'가 만드는 시한폭탄형 피해

공격자는 지금도 암호문을 수집할 수 있다. 문제는 그것이 '언제' 해독되느냐이다. 장기 보관 가치가 있는 데이터(국가 핵심기술 데이터, 연구 데이터, 의료정보, 장기 계약/법적 증거 등)는 5년, 10년 뒤에도 비밀이어야 한다. 따라서 양자위협은 '미래의 해킹'이 아니라 '오늘의 저장'이 만드는 시간차 위험이다.

양자컴퓨터 보안위협 대응을 위해 조직은 (1) 데이터의 기밀 유지 기간, (2) 시스템의 교체 · 업그레이드 주기, (3) 외부 의존(벤더 · 클라우드 · 표준)의 변화를 동시에 보며 양자보안 전환 시점을 앞당겨야 한다.

양자보안(Quantum Security) 쌍두마차: PQC와 QKD, 어떻게 바라볼 것인가?

'양자보안'의 정확한 의미: '디지털 데이터' 지키기

양자보안의 궁극적 목표는 미래 양자컴퓨터가 다루는 '큐비트 데이터'를 보호하는 것이 아니라, 현재 우리가 운영하는 디지털 시스템(비트 데이터)을 양자컴퓨터 공격에도 안전하게 만드는 것이다. 즉, 데이터는 '비트 데이터' 그대로이고 보호 수단(암호/키 관리/신뢰 인프라)이 '양자내성' 방향으로 바뀌는 것이다. 따라서 양자보안 전환은 양자컴퓨터의 공격으로부터 디지털 데이터를 지키기 위한 HW · FW · SW 전환과 제도 · 인식 · 장비 등의 전환을 모두 포괄하는 의미를 가진다.

PQC(Post-Quantum Cryptography, 양자내성암호): 디지털 기술 중심의 대규모 전환 핵심기술

PQC는 양자 공격(대용량 양자컴퓨터와 특수 양자 알고리즘 기반 암호해독 공격)까지 고려하여 이에 대응할 수 있는 수학적 난제를 기반으로 한 다양한 공개키 암호 알고리즘을 통칭한다. 가장 현실적인 강점은 기존 인프라(소프트웨어 · 펌웨어 · 하드웨어 등) 업데이트로 확장할 수 있다는 점이며, 표준화된 알고리즘을 라이브러리 · 프로토콜 · 제품에 탑재하는 '익숙한 설계와 구현'으로 전환이 가능하다는 점이다.

- 장점: 기존 디지털 기기 업데이트, 인터넷 규모 확장, 표준 기반 상호운용성, 단계적 탑재 가능.
- 주의점: 키 · 서명 크기 증가, 시스템 성능 저하, 기존 환경과의 호환성 이슈, 지속적 취약점 관리 필요.

QKD(Quantum Key Distribution, 양자 키 분배): 양자 기술 기반 Peer-to-Peer '키 전달'을 강화

QKD는 광자 등 양자 상태를 이용해 통신 양단이 동일한 키를 공유하고, 도청 시도를 물리적으로 탐지할 수 있다는 아이디어에 기반한다. QKD 구현을 위해서는 양자 상태를 생성하기 위한 특수한 양자 생성 · 송수신 장치가 필요하다. 실제 구축에서는 양자 송수신을 담당하는 특수 장치를 통해 만든 키를 디지털 서버인 QKMS에서 관리하고, 데이터 구간은 AES/ARIA/OTP 등 디지털 대칭키 알고리즘으로 암호화하는 구조가 일반적이다.

- **QKD(Quantum Key Distribution, 양자 키 분배) :** 양자 신호(예, 광자) 생성 · 검출과 양자 채널 기반 키 분배 프로토콜(BB84 알고리즘 등) 수행으로 양단에서 (비트열로 구성된) 키를 공유함. 특수하게 제작된 양자 생성 · 송수신 장치와 양자 채널에서 수행됨.
- **QKMS(Quantum Key Management System, 양자 키 관리 시스템) :** 키 저장 · 분배 · 수명 정책 · 감사 로그 관리 등의 역할을 수행함. 디지털 서버 시스템에서 수행됨. 즉, QKD를 통해 생성된 비트열로 구성된 키를 관리하는 서버 시스템을 의미함.
- **QENC(Quantum Encryption, 양자 통신 암호화) :** 생성 키를 이용해 기존 디지털 통신망에서 대칭키 기반 암복호화를 수행함. 디지털 기기에서 수행됨. 즉, QKMS가 관리하는 키를 제공받아 디지털 정보를 암복호화하는 기능을 의미함.

표 5-1 PQC와 QKD 비교(요약)

구분	PQC	QKD
기술적 배경	수학 문제(양자 공격 고려)	양자 물리(도청 탐지, 키 생성)
적용 방안	SW/프로토콜 업데이트 중심	특수 장비 · 광링크 · 운영 체계 필요
확장성	인터넷 규모 확장 용이	구간 중심(백본/전용망)에서 유리
주요 한계점	구현 · 성능 · 취약점 검증 등	비용 · 거리 · 운영 · 단말 보안 등

최신 기술 동향: '표준화–프로토콜 탑재–운영 체계'가 동시에 움직인다.

표준화의 분기점: NIST의 ML–KEM, ML–DSA, SLH–DSA (2024년)

PQC 전환이 본격화된 신호는 NIST가 발행한 표준이다. ML–KEM(격자 기반 키 교환, FIPS 203, 2024년 8월 발행), ML–DSA(격자 기반 전자서명, FIPS 204, 2024년 8월 발행), SLH–DSA(해시 기반 전자서명, FIPS 205, 2024년 8월 발행), FN–DSA(격자 기반 전자서명, 2026년에 FIPS 206으로 발행 예정)는 조직이 '무엇을, 어디에, 어떤 방식으로 탑재할까'를 논의할 수 있는 핵심 알고리즘을 제공한다.

국내에서는 KpqC 연구단이 2025년 1월에 4개의 국산 PQC 알고리즘(키 교환용 NTRU+, SMAUG–T 및 전자서명용 HAETAE, AIMer)를 선정했으며, 현재 표준화 작업을 추진 중이다.

PQC 전환에 있어서, 특히 전자서명 전환은 공급망 보안과도 직결된다. 배포 · 업데이트 · 서명이 무너지면 공격자는 '정상 업데이트'로 위장한 침투가 가능해진다. 따라서 코드서명 · OTA 검증자(단말/에이전트) 업데이트를 포함한 End–to–End PQC 전환 설계가 필요하다.

프로토콜이 바뀐다: TLS 1.3과 SSH의 '하이브리드(기존 + 양자내성)' 키 합의

디지털 데이터 보호를 위한 양자보안 전환에서는 프로토콜 협상이 매우 중요하다. 최근 흐름은 기존 ECDHE와 PQC–KEM을 결합한 하이브리드 키 합의를 우선 적용해, 현재 디지털 안전성과 양자 안전성을 동시에 확보하는 것이다.

- TLS 1.3: (예) X25519 + ML–KEM–768 결합으로 세션키 도출.
- SSH: 운영 접속/자동화 접속에서 하이브리드 Key Exchange 활용 확산.
- 현장 이슈: 핸드셰이크 메시지 증가 → 최대 데이터 패킷 크기 및 중간장비 호환성 사전 검증 필수.

QKD의 '네트워크화': QKMS, API 표준, 그리고 검증 체계

QKD는 단일 링크를 넘어 네트워크로 키를 공급하는 인프라로 진화하고 있다. 이 과정에서 핵심은 QKMS와 응용 간 인터페이스 표준이며, ETSI의 QKD API 규격은 상호운용성을 높여 벤더 종속을 낮추는 방향을 제시한다.

또한 공공 · 국가망 환경에서는 도입과 운영 사이에 보안성 검증의 절차가 존재한다. QKD 기반의 양자암호통신 장비들도 보안 기능이 탑재된 제품으로서 운용 전 국정원이 규정한 보안적합성 검증을 통과하여 보안기능확인서를 발급 받아야 하며, 경우에 따라서는 취약점 제거가 요구될 수 있다.

장점과 단점, 그리고 흔한 오해: PQC vs QKD, '정답'이 있을까?

PQC: 디지털 전환에 적합하지만, '지속적 취약점 관리'가 승부처

PQC는 가장 현실적인 대규모 전환 경로지만, 향후 양자컴퓨터와 양자 알고리즘의 발전 및 구현 취약점이 발목을 잡을 수 있다. NIST 표준으로 제정된 PQC 알고리즘과 국내 KpqC 선정 알고리즘들은 양자컴퓨터가 있더라도 아직까지는 효율적인 해법이 알려져 있지 않다는 전제 하에 개발된 것이다. 따라서 현재 PQC 알고리즘들의 수학적 난제를 풀 수 있는 효율적인 양자 알고리즘과 양자컴퓨터의 등장 여부를 지속적으로 모니터링하면서 취약점 관리를 유지해야 한다. 또한 디지털 기기에 구현할 때는 난수 품질, 상수시간 구현, 메모리 안전성, 예외처리, 부채널 공격 가능성 여부 등을 체계적으로 점검해야 한다.

- 성능/크기: 키 · 서명 크기 증가, 핸드셰이크 패킷 증가, CPU 부담 증가 가능.
- 구현 취약점: 부채널 분석 공격(타이밍/캐시/전력/기타), 라이브러리 버그, 저품질 난수 등.
- 운영 이슈: 알고리즘 교체가 반복될 수 있으므로 '암호화 민첩성(Crypto-Agility)'가 필수.

QKD: 양자 기술 기반의 '키 전달'은 강해지지만 엔드포인트를 대신하지 않는다.

QKD는 키 교환 과정의 도청 탐지 기능을 제공하지만, 키 교환을 수행하는 상대방이 정당한 권한을 가진 진짜 상대방인지 확인하기 어렵고 단말이 침해되면 키가 유출될 수 있는 단점이 있다. 또한 장거리 · 다자 확장에는 비용과 운용 제약이 존재한다. 따라서 QKD는 모든 구간의 대체재가 아니라 '핵심 구간'에서 키 공급 인프라로 결합될 때 효과가 크다.

보안 효과 극대화를 위한 '하이브리드(PQC+QKD)': 구간 분리 + 키 결합(XOR)

양자컴퓨터 보안위협에 대응하는 측면에서 보면, PQC와 QKD는 경쟁이라기보다 협력 관계이다. 대표적으로 ① 백본은 QKD, 단말 · 서비스 구간은 PQC로 구간을 분리하거나, ② 각각 만든 키를 XOR 결합해 세션키를 만드는 방식, ③ 응용 계층 End-to-End는 PQC를 사용하고 내부 핵심 구간에서 QKD 키 공급을 혼합하는 방식이 있다.

1) 구간 분리: 백본 · 기관 간 전용 구간은 QKD, 서비스/단말 구간은 PQC 사용.
2) 키 결합: '세션키 = (PQC 키) XOR (QKD 키)'로 이중 안전성 확보.
3) 혼합 운영: End-to-End는 PQC, 백본은 QKD로 키를 공급.

양자보안 전환 로드맵: '진단-우선순위-이행'의 3단계 절차

양자보안 전환은 단발성 프로젝트가 아니라 '다년간 반복되는 전환'이다. 따라서 조직은 먼저 암호 자산을 식별하고, 데이터 수명과 시스템 수명을 함께 고려해 우선순위를 결정한 뒤, 하이브리드 적용부터 시작하여 점진적으로 완전한 양자내성 환경으로 확장하는 로드맵이 필요하다.

표 5-2 단계별 양자보안 전환 로드맵(예시)

단계	우선 대상	핵심 작업	기간(예시)
1	전사 공통	암호 자산 인벤토리 구축, 장기 기밀 데이터 식별, 대칭키/해시 키 길이 상향	즉시
2	대외 서비스/업무망	TLS/SSH 하이브리드 시험, PKI · 인증서 정비, 다운그레이드 방어	3개월
3	공급망/서명 체계	코드서명 · OTA · 펌웨어 전환 설계, 검증자 업데이트, 키 관리(HSM/KMS) 강화	3~12개월
4	고등급 구간/전용망	QKD 도입 검토, QKMS 연동, PQC-QKD 키 결합 및 장애 Failover	24개월
5	전사 운영체계	암호화 민첩성(정책 · 자동화 · 검증) 정착, 취약점 관리에 따른 반복 업그레이드	상시

양자보안 전환 전략: 기술과 운영이 함께

암호 자산 인벤토리(Crypto asset inventory)는 '전환의 지도'이다.

양자보안 전환의 첫 단계는 '어디에 어떤 암호가 숨어 있는지' 찾는 것이다. RSA/ECC는 인증서 파일뿐 아니라 코드, 장비 펌웨어, 백업 스크립트, 내부 API, 클라우드 서비스 설정 등 곳곳에 존재한다. 암호 자산 인벤토리 파악이 부실하면 전환 뒤에도 빈틈이 남기 때문에 암호 자산 파악은 양자보안 전환의 첫 걸음으로 매우 중요하다.

- PKI/인증서: 서버/클라이언트 인증서, 내부 CA, 자동 갱신을 포함한 인증서 관리 자동화, 인증서 체인 등.
- 키 교환/터널: TLS, mTLS, VPN(IKE/IPsec), SSH KEX, 메시징 암호화 등.
- 서명/공급망: 코드서명, 컨테이너 이미지 서명, OTA/펌웨어 서명 등.
- 저장/보관: 암호화 백업, 장기 보관 아카이브, 데이터 레이크, HSM/KMS 등

데이터 수명 × 시스템 수명: 우선순위는 '데이터 가치와 보호 기간'이 결정한다.

장기 보관 데이터는 '선 수집, 후 해독' 위험이 크다. 시스템 교체 주기가 빈번하더라도 보호 가치가 높은 데이터가 오래 남는 상황이면 지금부터 전환을 시작해야 한다. 이와는 달리 OT/임베디드처럼 교체가 어려운 환경은 전환 비용이 크므로, 데이터 가치 분석 후 표준/벤더 로드맵과 함께 현실적 시나리오를 마련해야 한다.

암호화 민첩성(Crypto-Agility): '다음 교체'를 쉽게 만들어야 한다.

새로운 PQC 알고리즘이 개발될 수 있고 구현 취약점이 발견될 수 있으며, 보안성 및 성능은 지속적으로 개선될 것으로 기대된다. 따라서 조직은 신속하고 안전하게 알고리즘을 반복적으로 교체할 수 있는 구조(정책 · 자동화 · 검증)를 운영 역량으로 내재화해야 한다.

- 정책 기반 구성: 알고리즘/키 길이/프로토콜 버전을 중앙 정책으로 정의.
- 자동 검증: 상호운용 · 성능 · 최대 데이터 패킷 크기 · 장애 시나리오 자동 검증 등.
- 단계적 배포: 하이브리드에서 점진적으로 완전한 양자내성 전환, 폴백 최소화, 다운그레이드 방어 등.
- 벤더 관리: 양자내성 지원 · 검증을 계약 요구사항에 반영.

QKD 도입 체크 포인트: '장비'보다 '운영 모델'

QKD는 키 정책과 운영 모델이 완성되어야 의미가 있다. 키 생성률과 소비률, 키 수명, 장애 우회, 감사 로그, 검증 대응까지 운영 시나리오를 면밀하게 분석하여 수립해야 한다. 또한 QKD는 상대 인증을 해결하지 못하므로, 별도의 신뢰 검증 기법과 결합 설계가 필요하다.

- 적용 구간: 전용 회선/기관 간 링크/백본 등부터 단계적 적용.
- QKMS 연동: 키 사용 · 폐기 · 할당 정책과 장애 Failover(예: PQC로 자동 전환 등).
- 감사/검증: 로그 · 취약점 관리 체계 확보.
- 결합: PQC–QKD 키 결합(XOR)으로 단일 실패 지점 축소.

제도 · 검증의 관점: 대규모 활용의 선순환, '표준–검증–조달' 연계

공공 · 규제 환경에서는 표준 준수와 검증, 조달 요건이 동시에 움직인다. QKD 기반의 양자암호통신 장비도 운용 전 보안적합성 검증과 취약점 제거가 요구될 수 있으며, PQC는 KCMVP 제도의 암호모듈 검증과 결합해 확산 경로가 형성될 가능성이 크다. 따라서 양자보안 전환을 위해서 조직은 지금부터 '검증 가능한 설계 및 구현'과 '증명 가능한 자료(로그/정책)' 중심 운영을 준비해야 한다.

분야별 적용 포인트: 웹 · 클라우드 · OT/IoT · 공급망에서 무엇이 달라지나?

양자보안 전환은 한 번에 이뤄지지 않는다. 조직은 디지털 데이터 및 암호 자산이 '가장 많이 노출되고, 가장 오래 남고, 가장 자주 쓰이는' 경로부터 바꿔야 한다. 특히 인터넷 대외 서비스, 클라우드 워크로드, OT/IoT 단말, 공급망(배포 · 업데이트)은 서로 다른 제약이 있으므로, 동일한 처방을 일괄 적용하면 오히려 장애와 보안 공백을 만들 수 있다. 따라서 각 영역별 특성에 맞는 전환 전략이 필요하다.

표 5-3 영역별 양자보안 전환 포인트(요약)

영역	주요 의존 지점	우선 전환 요소	주요 고려사항
웹/대외 서비스	TLS 키 교환, 서버 인증서	TLS 하이브리드 키 교환, 인증서 체인 준비	프록시/로드밸런서/보안장비 협상 실패 점검
클라우드/마이크로서비스	mTLS, 서비스 메시, API 게이트웨이	라이브러리/사이드카의 PQC 지원, 정책 기반 배포	구성 드리프트 탐지 · 자동 롤백
운영/접속(SSH)	KEX · 호스트키 · 자동화 계정	SSH 하이브리드 KEX, 호스트키 정책	배포/백업 자동화부터 우선 전환
공급망/업데이트	코드서명, OTA 검증	PQC 서명 도입, 검증자 업데이트	부트체인 · 타임스탬프 연동 고려
OT/IoT	장시간 사용 펌웨어, 제한 자원	경량 전환 경로, 게이트웨이 기반 보호	현장 교체 주기 · 벤더 로드맵 확보
장기 보관	아카이브 암호화 · 서명	재암호화 · 재서명 계획	데이터 수명 기반 우선순위화

여러 서비스 영역에서 공통적으로 자주 놓치는 3가지는 다음과 같으며, 이러한 부분을 염두에 두고 양자보안 전환 전략을 수립해야 한다.

① '폴백이 보안 구멍'이 된다: 호환성 때문에 마련한 폴백 경로가 다운그레이드 공격의 통로가 될 수 있으므로, 폴백을 최소화하고 정책적으로 차단해야 한다.
② '검증자가 더 중요'하다: 서명 전환은 서명자(빌드 서버)보다 검증자(단말/에이전트)가 업데이트되어야 완성된다.
③ '키 관리 없이는 여전히 취약'하다: HSM/KMS, 키 생명주기 관리 정책, 감사 로그가 없으면 알고리즘만 바꿔도 위험이 남는다.

'계획'이 아니라 '실행'이 핵심이다

양자보안은 '양자컴퓨터의 등장 시점'을 맞히는 문제가 아니다. 현재의 공개키 기반 신뢰 인프라가 양자 기술에 구조적으로 취약하다는 사실은 변하지 않는다. 따라서 양자보안 전환의 목표는 "언제 깨질지 예측"이 아니라 "깨지기 전에 옮겨갈 수 있는 길(전환 능력)을 확보"하는 것이다.

PQC는 인터넷 규모로 확장 가능한 현실적 경로를 제공하고, QKD는 특정 구간에서 키 전달 인프라를 보강한다. 둘 중 하나를 고르는 것이 아니라, 데이터 수명과 시스템 수명, 운용 제약, 가용 예산을 기준으로 '구간별 하이브리드 전략'을 설계하는 것이 현명한 접근이라 할 수 있다.

지금 해야 할 일은 명확하다. 암호 자산 인벤토리를 만들고, 암호화 민첩성 기반 설계를 시작하며, 서명 · 공급망 영역을 우선 강화하는 기술 적용을 실행에 옮기는 것이다. 양자보안은 미래의 과제가 아니라, 오늘 수집되는 디지털 데이터가 내일의 침해가 되지 않게 만드는 현재형 과제다.

참고문헌 (주요 표준/가이드라인)

- NIST, PQC 표준 (FIPS 203(ML-KEM), FIPS 204(ML-DSA), FIPS 205(SLH-DSA)).
- IETF, draft-ietf-tls-ecdhe-mlkem (TLS 1.3 하이브리드 키 합의).
- UK NCSC, Timelines for migration to post quantum cryptography.
- Google Chromium Blog, PQC-TLS 기술.
  https://blog.chromium.org/2023/08/protecting-chrome-traffic-with-hybrid.html

부록 양자보안 전환 체크리스트(참조용)

양자보안은 '알고리즘 교체'로 끝나지 않는다. 조직이 실제로 움직이기 위해서는 (1) 누가(조직/역할) 무엇(자산/정책)을 언제(마일스톤)까지 바꾸는지, 그리고 (2) 무엇을 남길지(전환 관련 로그)가 문서로 정리되어야 한다. 아래 체크리스트는 표준 · 프로토콜 변화가 빠른 상황을 고려하여 운영 관점에서 양자보안 전환을 시작하기 위한 '참조용' 도구다.

표 5-4 양자보안 전환 체크리스트(운영 관점)

구분	핵심 질문	권장 조치(예시)	산출물/로그
데이터	장기 기밀 데이터는 무엇인가?	데이터 분류(보호기간) 재정의, '선 수집, 후 해독' 위험 등급 부여	자산 목록, 분류 기준, 보관 기간
암호 자산	RSA/ECC 사용 지점은 어디인가?	인증서/키 교환/서명/백업까지 탐색 자동화	Crypto inventory, CBOM/SBOM/HBOM 연계
TLS/SSH	하이브리드 전환을 어디부터 시작할까?	대외 서비스→업무망→내부 TLS 순서로 전환	테스트 결과, 호환성/최대 데이터 패킷 크기 리포트
서명/공급망	검증자(단말)는 업데이트 가능한가?	코드서명 전환 설계 + 검증자 업데이트 계획	전환 설계서, 릴리즈/롤백 플랜
키 관리	키는 어디서 생성/보관/회전되나?	HSM/KMS 연동, 키 생명주기 관리 정책과 감사 로그 강화	키 정책, 접근통제, 감사 로그
운영	알고리즘 교체를 반복할 수 있나?	Crypto-Agility(정책/자동화/검증) 체계화	정책 템플릿, 자동화 파이프라인
QKD(선택)	적용 구간과 운영 모델은?	백본/전용 링크 중심 검토, Failover(PQC로 대체) 설계	구간 설계, 키 소비/생성 분석, 운영절차

하반기

25년 사이버 위협 동향 및 26년 전망

KISA 한국인터넷진흥원

[나주본원] 전라남도 나주시 진흥길 9 한국인터넷진흥원

[서울청사] 서울시 송파구 중대로 135 (가락동) IT벤처타워 TEL. 1544-5118 / 02-405-5118